



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/574, 33/48, C12Q 1/68		A1	(11) International Publication Number: WO 99/54738 (43) International Publication Date: 28 October 1999 (28.10.99)
(21) International Application Number: PCT/US99/05766 (22) International Filing Date: 16 March 1999 (16.03.99)		(81) Designated States: AU, CA, CN, JP, KR, NZ, ZA, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority Data: 09/061,709 17 April 1998 (17.04.98) US		Published <i>With international search report.</i>	
(71) Applicant: LUDWIG INSTITUTE FOR CANCER RESEARCH [CH/US]; 605 Third Avenue, New York, NY 10103 (US).			
(72) Inventors: CHEN, Yao-Tseng; 525 East 68 Street, New York, NY 10021 (US). GURE, Ali; 1275 York Avenue, New York, NY 10021 (US). TSANG, Solam; 1275 York Avenue, New York, NY 10021 (US). STOCKERT, Elisabeth; 1275 York Avenue, New York, NY 10002 (US). JAGER, Elke; Steinbacher Hohl 2-28, D-60488 Frankfurt am Main (DE). KNUTH, Alexander; Steinbacher Hohl 2-28, D-60488 Frankfurt am Main (DE). OLD, Lloyd, J.; 605 Third Avenue, New York, NY 10021 (US).			
(74) Agent: HANSON, Norman, D.; Fulbright & Jaworski LLP, 666 Fifth Avenue, New York, NY 10103 (US).			
(54) Title: ISOLATED NUCLEIC ACID MOLECULE ENCODING CANCER ASSOCIATED ANTIGENS, THE ANTIGENS PER SE, AND USES THEREOF			
(57) Abstract The invention relates to newly identified cancer associated antigens, referred to as CT7, KOC-2 and KOC-3. The invention also relates to observations regarding known molecule KOC-1. It has been discovered that each of these molecules provokes antibodies when expressed by a subject. The ramifications of this observation are also a part of this invention.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

ISOLATED NUCLEIC ACID MOLECULE ENCODING
CANCER ASSOCIATED ANTIGENS, THE ANTIGENS PER SE,
AND USES THEREOF

RELATED APPLICATION

This application is a continuation in part of Serial No. 09/061,709 filed April 17, 1998, incorporated by reference.

FIELD OF THE INVENTION

This invention relates to antigens associated with cancer, the nucleic acid molecules encoding them, as well as the uses of these.

BACKGROUND AND PRIOR ART

It is fairly well established that many pathological conditions, such as infections, cancer, autoimmune disorders, etc., are characterized by the inappropriate expression of certain molecules. These molecules thus serve as "markers" for a particular pathological or abnormal condition. Apart from their use as diagnostic "targets", i.e., materials to be identified to diagnose these abnormal conditions, the molecules serve as reagents which can be used to generate diagnostic and/or therapeutic agents. A by no means limiting example of this is the use of cancer markers to produce antibodies specific to a particular marker. Yet another non-limiting example is the use of a peptide which complexes with an MHC molecule, to generate cytolytic T cells against abnormal cells.

Preparation of such materials, of course, presupposes a source of the reagents used to generate these. Purification from cells is one laborious, far from sure method of doing so. Another preferred method is the isolation of nucleic acid molecules which encode a particular marker, followed by the use of the isolated encoding molecule to express the desired molecule.

Two basic strategies have been employed for the detection of such antigens, in e.g., human tumors. These will be referred to as the genetic approach and the biochemical approach. The genetic approach is exemplified by, e.g., dePlaen et al., Proc. Natl. Sci. USA 85: 2275

(1988), incorporated by reference. In this approach, several hundred pools of plasmids of a cDNA library obtained from a tumor are transfected into recipient cells, such as COS cells, or into antigen-negative variants of tumor cell lines which are tested for the expression of the specific antigen. The biochemical approach, exemplified by, e.g., O. Mandelboim, et al., *Nature* 369: 69 (1994) incorporated by reference, is based on acidic elution of peptides which have bound to MHC-class I molecules of tumor cells, followed by reversed-phase high performance liquid chromatography (HPLC). Antigenic peptides are identified after they bind to empty MHC-class I molecules of mutant cell lines, defective in antigen processing, and induce specific reactions with cytotoxic T-lymphocytes. These reactions include induction of CTL proliferation, TNF release, and lysis of target cells, measurable in an MTT assay, or a ⁵¹Cr release assay.

These two approaches to the molecular definition of antigens have the following disadvantages: first, they are enormously cumbersome, time-consuming and expensive; and second, they depend on the establishment of cytotoxic T cell lines (CTLs) with predefined specificity.

The problems inherent to the two known approaches for the identification and molecular definition of antigens is best demonstrated by the fact that both methods have, so far, succeeded in defining only very few new antigens in human tumors. See, e.g., van der Bruggen et al., *Science* 254: 1643-1647 (1991); Brichard et al., *J. Exp. Med.* 178: 489-495 (1993); Coulie, et al., *J. Exp. Med.* 180: 35-42 (1994); Kawakami, et al., *Proc. Natl. Acad. Sci. USA* 91: 3515-3519 (1994).

Further, the methodologies described rely on the availability of established, permanent cell lines of the cancer type under consideration. It is very difficult to establish cell lines from certain cancer types, as is shown by, e.g., Oettgen, et al., *Immunol. Allerg. Clin. North. Am.*

10: 607-637 (1990). It is also known that some epithelial cell type cancers are poorly susceptible to CTLs in vitro, precluding routine analysis. These problems have stimulated the art to develop additional methodologies for identifying cancer associated antigens.

One key methodology is described by Sahin, et al., Proc. Natl. Acad. Sci. USA 92: 11810-11813 (1995), incorporated by reference. Also, see U.S. Patent No. 5,698,396, and Application Serial No. 08/479,328, filed on June 7, 1995 and January 3, 1996, respectively. All three of these references are incorporated by reference. To summarize, the method involves the expression of cDNA libraries in a prokaryotic host. (The libraries are secured from a tumor sample). The expressed libraries are then immunoscreened with absorbed and diluted sera, in order to detect those antigens which elicit high titer humoral responses. This methodology is known as the SEREX method ("Serological identification of antigens by Recombinant Expression Cloning"). The methodology has been employed to confirm expression of previously identified tumor associated antigens, as well as to detect new ones. See the above referenced patent applications and Sahin, et al., supra, as well as Crew, et al., EMBO J 14: 2333-2340 (1995).

This methodology has been applied to a range of tumor types, including those described by Sahin et al., supra, and Pfreundschuh, supra, as well as to esophageal cancer (Chen et al., Proc. Natl. Acad. Sci. USA 94: 1914-1918 (1997)); lung cancer (Güre et al., Cancer Res. 58: 1034-1041 (1998)); colon cancer (Serial No. 08/948,705 filed October 10, 1997) incorporated by reference, and so forth. Among the antigens identified via SEREX are the SSX2 molecule (Sahin et al., Proc. Natl. Acad. Sci. USA 92: 11810-11813 (1995); Tureci et al., Cancer Res. 56: 4766-4772 (1996); NY-ESO-1 Chen, et al., Proc. Natl. Acad. Sci. USA 94: 1914-1918 (1997); and SCP1 (Serial No. 08/892,705 filed July 15, 1997) incorporated by reference.

Analysis of SEREX identified antigens has shown overlap between SEREX defined and CTL defined antigens. MAGE-1, tyrosinase, and NY-ESO-1 have all been shown to be recognized by patient antibodies as well as CTLs, showing that humoral and cell mediated responses do act in concert.

It is clear from this summary that identification of relevant antigens via SEREX is a desirable aim. The inventors have modified standard SEREX protocols and have screened a cell line known to be a good source of the antigens listed *supra*, using allogeneic patient sample. New antigens have been identified in this way and have been studied. Also, a previously known molecule has now been identified via SEREX techniques.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Example 1

The melanoma cell referred to as SK-MEL-37 was used, because it has been shown to express a number of members of the CT antigen family, including MAGE-1 (Chen et al., Proc. Natl. Acad. Sci. USA 91: 1004-1008 (1994); NY-ESO-1 (Chen et al. Proc. Natl. Acad. Sci. USA 94: 1914-1918 (1997)); and various members of the SSX family (Gure et al., Int. J. Cancer 72: 965-971 (1997)).

Total RNA was extracted from cultured samples of SK-MEL-37 using standard methods, and this was then used to construct a cDNA library in commercially available, λZAP expression vector, following protocols provided by the manufacturer. The cDNA was then transfected into E. coli and screened, following Sahin et al., Proc. Natl. Acad. Sci. USA 92: 11810-11813 (1995), incorporated by reference, and Pfreundschuh, U.S. Patent No. 5,698,396, also incorporated by reference. The screening was done with allogeneic patient serum "NW38." This serum had been shown, previously, to contain high titer antibodies against MAGE-1 and NY-

ESO-1. See, e.g., Jäger et al., *J. Exp. Med.* 187: 265-270 (1998), incorporated by reference. In brief, serum was diluted 1:10, preabsorbed with lysates of transfected *E. coli*, further diluted to 1:2000, and then incubated overnight at room temperature with nitrocellulose membranes containing phage plaques, prepared in accordance with Sahin et al., and Pfreundschuh, supra. The library contained a total of 2.3×10^7 primary clones. After washing, the filters were incubated with alkaline phosphatase conjugated, goat anti-human Fcγ secondary antibodies, and were then visualized by incubating with 5-bromo-4-chloro-3-indolyl phosphate, and nitroblue tetrazolium.

After screening 1.5×10^5 of the clones, a total of sixty-one positives had been identified. Given this number, screening was stopped, and the positive clones were subjected to further analysis.

Example 2

The positive clones identified in example 1, supra, were purified, the inserts were excised in vitro, and inserted into a commercially available plasmid, pBK-CMV, and then evaluated on the basis of restriction mapping with EcoRI and XbaI. Clones which represented different inserts on the basis of this step were sequenced, using standard methodologies.

There was a group of 10 clones, which could not be classified other than as "miscellaneous genes", in that they did not seem to belong to any particular family. They consisted of 9 distinct genes, of which four were known, and five were new. The fifty one remaining clones were classified into four groups. The data are presented in Tables 1 and 2, which follow.

The largest group are genes related to KOC ("KH-domain containing gene, overexpressed in cancer" which has been shown to be overexpressed in pancreatic cancer, and maps to

chromosome 7p11.5. See Müller-Pillasch et al., *Oncogene* 14: 2729-2733 (1997). Two of the 33 were derived from the KOC gene, and the other 31 were derived from two previously unidentified, but related genes. Examples 6 et seq. describe work on this group of clones.

Eleven clones, i.e., Group 2, were MAGE sequences. Four were derived from MAGE-4a, taught by DePlaen et al., *Immunogenetics* 40: 360-369, Genbank U10687, while the other 7 hybridized to a MAGE-4a probe, derived from the 5' sequence, suggesting they belong to the MAGE family.

The third group consisted of five clones of the NY-ESO-1 family. Two were identical to the gene described by Chen et al., *Proc. Natl. Acad. Sci. USA* 94: 1914-1918 (1997), and in Serial No. 08/725,182, filed October 3, 1996, incorporated by reference. The other three were derived from a second member of the NY-ESO-1 family, i.e., LAGE-1. See U.S. application Serial No. 08/791,495, filed January 27, 1997 and incorporated by reference.

The fourth, and final group, related to a novel gene referred to as CT7. This gene, the sequence of which is presented as SEQ ID NO: 1, was studied further.

Table 1. SEREX-identified genes from allogeneic screening of SK-MEL-37 library

Gene group	# of clones	Comments
KOC	33	derived from 3 related genes
MAGE	11	predominantly MAGE-4a (see text)
NY-ESO-1	5	derived from 2 related genes (NY-ESO-1, LAGE-1)
CT7	2	new cancer/testis antigen
Miscellaneous	10	see Table 2

Table 2. SEREX-identified genes from allogeneic screening of SK-MEL-37 library--
Miscellaneous group

Clone designation	Gene
MNW-4, MNW-7	S-adenyl homocysteine hydrolase
MNE-6a	Glutathione synthetase
MNW-24	proliferation-associated protein p38-2G4
MNW-27a	phosphoribosyl pyrophosphate synthetase-associated protein 39
MNW-6b	unknown gene, identical to sequence tags from pancreas, uterus etc.
MNW-14b	unknown gene, identical to sequence tags from lung, brain, fibroblast etc.
MNW-34a	unknown gene, identical to sequence tags from multiple tissues
MNW-17	unknown gene, identical to sequence tags from pancreas and fetus
MNW-29a	unknown gene, no significant sequence homology, universally expressed

Example 3

The two clones for CT7, referred to supra, were 2184 and 1965 base pairs long. Analysis of the longer one was carried out. It presented an open reading frame of 543 amino acids, which extended to the 5' end of the sequence, indicating that it was a partial cDNA clone.

In order to identify the complete sequence, and to try to identify additional, related genes, a human testicular cDNA library was prepared, following standard methods, and screened with probes derived from the longer sequence, following standard methods.

Eleven positives were detected, and sequenced, and it was found that all derived from the same gene. When the polyA tail was excluded, full length transcript, as per SEQ ID NO: 1, consisted of 4265 nucleotides, broken down into 286 base pairs of untranslated 5' - region, a coding region of 3429 base pairs, and 550 base pairs of untranslated 3' region. The predicted protein is 1142 amino acids long, and has a calculated molecular mass of about 125 kilodaltons.

See SEQ ID NO: 2.

The nucleotide and deduced amino acid sequences were screened against known databases, and there was some homology with the MAGE-10 gene, described by DePlaen et al., Immunogenetics 40: 360-369 (1994). The homology was limited to about 210 carboxy terminal amino acids, i.e., amino acids 908-1115 of the subject sequence, and 134-342 of MAGE-10. The percent homology was 56%, rising to 75% when conservative changes are included.

There was also extensive homology with a sequence reported by Lucas et al., Canc. Res. 58: 743-752 (1998), and application Serial No. 08/845,528 filed April 25, 1997, also incorporated by reference. A total of 14 nucleotides differ in the open reading frame, resulting in a total of 11 amino acids which differ between the sequences.

The 5' region of the nucleotide and sequence and corresponding amino acid sequence demonstrates a strikingly repetitive pattern, with repeats rich in serine, proline, glutamine, and leucine, with an almost invariable core of PQSPLQI (SEQ ID NO: 3). In the middle of the molecule, 11 almost exact repeats of 35 amino acids were observed. The repetitive portions make up about 70% of the entire sequence, begin shortly after translation initiation, at position 15, and ending shortly before the region homologous to MAGE 4a.

Example 4

The expression pattern for mRNA of CT7 was then studied, in both normal and malignant tissues. RT-PCR was used, employing primers specific for the gene. The estimated melting temperature of the primers was 65-70°C, and they were designed to amplify 300-600 base pair segments. A total of 35 amplification cycles were carried out, at an annealing temperature of 60°C. Table 3, which follows, presents the data for human tumor tissues. CT7 was expressed in a number of different samples. Of fourteen normal tissues tested, there was strong expression in testis, and none in colon, brain, adrenal, lung, breast, pancreas, prostate, thymus or uterus tissue. There was low level expression in liver, kidney, placenta and fetal brain, with fetal brain showing three transcripts of different size. The level of expression was at least 20-50 times lower than in testis. Melanoma cell lines were also screened. Of these 7 of the 12 tested showed strong expression, and one showed weak expression.

Table 3. CT7 mRNA expression in various humor tumors by RT-PCR

Tumor type	mRNA, positive/total
Melanoma	7/10
Breast cancer	3/10
Lung cancer	3/9
Head/neck cancer	5/14
Bladder cancer	4/9
Colon cancer	1/10
Leimyosarcoma	1/4
synovial sarcoma	2/4
Total	26/70

Example 5

Southern blotting experiments were then carried out to determine if CT7 belonged to a family of genes. In these experiments, genomic DNA was extracted from normal human tissues. It was digested with BamHI, EcoRI, and HindIII, separated on a 0.7% agarose gel, blotted onto a nitrocellulose filter, and hybridized, at high stringency (65°C, aqueous buffer), with a ³²P labelled probe, derived from SEQ ID NO: 1.

The blotting showed anywhere from two to four bands, suggesting one or two genes in the family.

Example 6

As noted in example 2, supra, thirty three of the sixty one positive clones were related to KOC. Clones were sequenced using standard methodologies. As indicated supra, one clone

was identical to KOC, initially reported by Müller-Pillasch, et al., supra. Given that two additional related sequences were identified, the known KOC gene is referred to as KOC-1 hereafter (SEQ ID NO: 4). The second clone, referred to as KOC-2 hereafter, was found once. The sequence is presented as SEQ ID NO: 5. Its deduced amino acid sequence is 72.5% identical to that for KOC-1.

The third sequence, KOC-3, appeared thirty times (SEQ ID NO: 6). Its deduced amino acid sequence is 63% identical to KOC-1.

Testicular cDNA libraries were analyzed in the same way that the SK-MEL-37 library was analyzed, i.e., with allogeneic serum from NW-38. See example 3, supra.

Following analysis of testicular libraries, a longer form of KOC-2 was isolated. This is presented as SEQ ID NO: 7. When SEQ ID NOS: 5 & 7 are compared, the former is 1705 base pairs in length, without a polyA tail. It contains 1362 base pairs of coding sequence, and 343 base pairs of 3' untranslated sequence. Nucleotides 275-1942 of SEQ ID NO: 7 are identical to nucleotides 38-1705 of SEQ ID NO: 5.

The sequence of KOC-3, set forth as SEQ ID NO: 6, is 3412 base pairs long, and consists of 72 base pairs of 5' untranslated region, 1707 base pairs of open reading frame, and 1543 base pairs of untranslated, 3' region. An alternate form was also isolated, (SEQ ID NO: 8), and is 129 base pairs shorter than SEQ ID NO: 6.

Example 7

Expression patterns for KOC-1, KOC-2 and KOC-3 were then studied, using RT-PCR and the following primer pairs:

GAAAGTATCT TCAAGGACGC C

CTGCAAGGGG TTTGCTGGG CG

(SEQ ID NOS: 9 & 10).

TCCTTGCACG CTGCAGCCTC AG

CCAAGTGGTG GCCATTAGCT TC

(SEQ ID NOS: 11 & 12)

GCTCTTGAGG GACAGGAAGG TC

GACGTTGACA ACGGCGGTTT CT

(SEQ ID NOS: 13 & 14).

SEQ ID NOS: 9 & 10 were designed to amplify KOC-1 while SEQ ID NOS: 11 & 12 were designed to amplify KOC-2, and SEQ ID NOS: 13 & 14 were designed to amplify KOC-3.

To carry out the RT-PCR, relevant primer pairs were added to cDNA samples prepared from various mRNAs by reverse transcription. PCR was then carried out at an annealing temperature of 60°C, and extension at 72°C, for 35 cycles. The resulting products were then analyzed by gel electrophoresis.

SEQ ID NOS 9 & 10 amplify nucleotides 305-748 of SEQ ID NO: 1. A variety of normal and malignant cell types were tested. Strong expression was found in testis, moderate expression in normal brain, and low levels of expression were found in normal colon, kidney, and liver.

The Müller-Pillasch paper, cited supra, identified expression of KOC-1 in pancreatic tumor cell lines, gastric cancer, and normal placenta, via Northern blotting. This paper also reported that normal heart, brain, lung, liver, kidney and pancreatic tissue were negative for

KOC-1 expression. The difference in results suggests that the level of expression of KOC-1 is very low in normal tissues.

When KOC-2 expression was studied, the only positive normal tissue was testis (brain, liver, kidney and colon were negative).

Modification of the protocol for detecting KOC-2 resulted in positives in normal kidney, liver and melanoma.

When KOC-3 expression was studied, it was found that the gene was universally expressed in normal tissues, with highest expression in testis.

The pattern of expression of KOC-3 in different melanoma cell lines was analyzed, using standard Northern blotting. Over expression in several cell lines was observed, which is consistent with the more frequent isolation of this clone than any other.

Example 8

A study was carried out to determine if KOC-1 is expressed at higher levels in melanoma cells, as compared to normal skin cells. This was done using representational difference analysis, or "RDA." See Lisitsyn, et al. *Science* 259: 946-951 (1993), and O'Neill, et al. *Nucl. Acids Res.* 25:2681-2 (1997), both of which are incorporated by reference. Specifically, tester cDNA was taken from SK-MEL-37, and driver cDNA was taken from a skin sample representing mRNA from various cell types in the skin. The cDNAs were digested with either Tsp509I, Hsp92II, or DpnII. When DpnII was the enzyme used for digestion, adaptor oligonucleotides R-Bgl-24, J-Bgl-24, and N-Bgl-24 described by O'Neill, et al., *supra*, and Hubank, et al. *Nucl. Acids Res.*

WO 99/54738

22:5640-5648 (1994) were used. When Tsp509I was the endonuclease, the same adaptors were used, as were R-Tsp-12, i.e.:

AATTTGCGGT GA

(SEQ ID NO: 15)

J-Tsp-12, i.e.:

AATTTGTTCA TG

(SEQ ID NO: 16)

and N-Tsp-12, i.e.:

AATTTCCCT CG

(SEQ ID NO: 17)

When Hsp92II was the endonuclease, the adaptors were:

R-Hsp-24, i.e.:

AGCACTCTCC AGCCTCTCAC CATG

(SEQ ID NO: 18);

J-Hsp-24, i.e.:

ACCGACGTCG ACTATCATG CATG

(SEQ ID NO: 19);

N-Hsp-24, i.e.:

AGGCAACTGT GCTATCCGAG CATG

(SEQ ID NO: 20);

R-Hsp-8, i.e.:

GTGAGAGG

(SEQ ID NO: 21);

J-Hsp-8, i.e.:

CATGGATG

(SEQ ID NO: 22);

N-Hsp-8, i.e.:

CTCGGATA

(SEQ ID NO: 23).

In order to hybridize tester and driver, either 3XEE buffer (30mM EPPS, pH8, 3mM EDTA), or a buffer of 2.4M tetraethylammonium chloride (TEACl) 3mM EDTA, 10mM Tris HCl, pH8, was used. When DNA was dissolved in 10 μ l of TEACl buffer, it was denatured at 80°C for 10 minutes, followed by renaturing at 42°C for 20 hours. Amplicons were gel purified, and the DP3 or DP2 product was ligated into BamHI (when DpnII was used), EcoRI (when Tsp 509I was used), or SpHI (when Hsp92II was used), cloning vectors were digested, and then sequenced. Sequence analysis of the cDNA molecules derived from these experiments identified KOC-1 as one of the genes isolated, indicating that KOC-1 mRNA is present at a higher level in Sk-Mel 37 cells as compared to normal skin cells.

The foregoing examples describe the isolation of a nucleic acid molecule which encodes a cancer associated antigen. "Associated" is used herein because while it is clear that the relevant molecule was expressed by several types of cancer, other cancers, not screened herein, may also express the antigen.

The invention relates to those nucleic acid molecules which encode the antigens CT7, KOC-2 and KOC-3, as described herein, such as a nucleic acid molecule consisting of the nucleotide sequence SEQ ID NO: 1, molecules comprising the nucleotide sequence of SEQ ID

NO: 5, 6, 7 or 8 and so forth. Also embraced are those molecules which are not identical to SEQ ID NOS: 1, 5, 6, 7 or 8, but which encode the same antigen.

Also a part of the invention are expression vectors which incorporate the nucleic acid molecules of the invention, in operable linkage (i.e., "operably linked") to a promoter. Construction of such vectors, such as viral (e.g., adenovirus or Vaccinia virus) or attenuated viral vectors is well within the skill of the art, as is the transformation or transfection of cells, to produce eukaryotic cell lines, or prokaryotic cell strains which encode the molecule of interest. Exemplary of the host cells which can be employed in this fashion are COS cells, CHO cells, yeast cells, insect cells (e.g., Spodoptera frugiperda), NIH 3T3 cells, and so forth. Prokaryotic cells, such as E. coli and other bacteria may also be used. Any of these cells can also be transformed or transfected with further nucleic acid molecules, such as those encoding cytokines, e.g., interleukins such as IL-2, 4, 6, or 12 or HLA or MHC molecules.

Also a part of the invention are the antigens described herein, both in original form and in any different post translational modified forms. The molecules are large enough to be antigenic without any posttranslational modification, and hence are useful as immunogens, when combined with an adjuvant (or without it), in both precursor and post-translationally modified forms. Antibodies produced using these antigens, both poly and monoclonal, are also a part of the invention as well as hybridomas which make monoclonal antibodies to the antigens. The whole protein can be used therapeutically, or in portions, as discussed infra. Also a part of the invention are antibodies against this antigen, be these polyclonal, monoclonal, reactive fragments, such as Fab, (F(ab)₂)' and other fragments, as well as chimeras, humanized antibodies, recombinantly produced antibodies, and so forth.

As is clear from the disclosure, one may use the proteins and nucleic acid molecules of the invention diagnostically. The SEREX methodology discussed herein is premised on an immune response to a pathology associated antigen. Hence, one may assay for the relevant pathology via, e.g., testing a body fluid sample of a subject, such as serum, for reactivity with the antigen per se. Reactivity would be deemed indicative of possible presence of the pathology. So, too, could one assay for the expression of any of the antigens via any of the standard nucleic acid hybridization assays which are well known to the art, and need not be elaborated upon herein. One could assay for antibodies against the subject molecules, using standard immunoassays as well.

Analysis of SEQ ID NO: 1, 5, 6, 7 and 8 will show that there are 5' and 3' non-coding regions presented therein. The invention relates to those isolated nucleic acid molecules which contain at least the coding segment, i.e., nucleotides 54-593, of SEQ ID NO: 1, nucleotides 1-1019 of SEQ ID NO: 3, nucleotides 73-1780 of SEQ ID NO: 8, and so forth, and which may contain any or all of the non-coding 5' and 3' portions.

Also a part of the invention are portions of the relevant nucleic acid molecules which can be used, for example, as oligonucleotide primers and/or probes, such as one or more of SEQ ID NOS: 7, 8, 9, 10, 11, 12, 13 or 14 as well as amplification product like nucleic acid molecules comprising at least nucleotides 305-748 of SEQ ID NO: 1.

As was discussed supra, study of other members of the "CT" family reveals that these are also processed to peptides which provoke lysis by cytolytic T cells. There has been a great deal of work on motifs for various MHC or HLA molecules, which is applicable here. Hence, a further aspect of the invention is a therapeutic method, wherein one or more peptides derived

from the antigens of the invention which bind to an HLA molecule on the surface of a patient's tumor cells are administered to the patient, in an amount sufficient for the peptides to bind to the MHC/HLA molecules, and provoke lysis by T cells. Any combination of peptides may be used. These peptides, which may be used alone or in combination, as well as the entire protein or immunoreactive portions thereof, may be administered to a subject in need thereof, using any of the standard types of administration, such as intravenous, intradermal, subcutaneous, oral, rectal, and transdermal administration. Standard pharmaceutical carriers, adjuvants, such as saponins, GM-CSF, and interleukins and so forth may also be used. Further, these peptides and proteins may be formulated into vaccines with the listed material, as may dendritic cells, or other cells which present relevant MHC/peptide complexes.

Similarly, the invention contemplates therapies wherein nucleic acid molecules which encode the proteins of the invention, one or more or peptides which are derived from these proteins are incorporated into a vector, such as a Vaccinia or adenovirus based vector, to render it transfectable into eukaryotic cells, such as human cells. Similarly, nucleic acid molecules which encode one or more of the peptides may be incorporated into these vectors, which are then the major constituent of nucleic acid bases therapies.

Any of these assays can also be used in progression/regression studies. One can monitor the course of abnormality involving expression of these antigens simply by monitoring levels of the protein, its expression, antibodies against it and so forth using any or all of the methods set forth supra.

It should be clear that these methodologies may also be used to track the efficacy of a therapeutic regime. Essentially, one can take a baseline value for a protein of interest using any

of the assays discussed supra, administer a given therapeutic agent, and then monitor levels of the protein thereafter, observing changes in antigen levels as indicia of the efficacy of the regime.

As was indicated supra, the invention involves, inter alia, the recognition of an "integrated" immune response to the molecules of the invention. One ramification of this is the ability to monitor the course of cancer therapy. In this method, which is a part of the invention, a subject in need of the therapy receives a vaccination of a type described herein. Such a vaccination results, e.g., in a T cell response against cells presenting HLA/peptide complexes on their cells. The response also includes an antibody response, possibly a result of the release of antibody provoking proteins via the lysis of cells by the T cells. Hence, one can monitor the effect of a vaccine, by monitoring an antibody response. As is indicated, supra, an increase in antibody titer may be taken as an indicia of progress with a vaccine, and vice versa. Hence, a further aspect of the invention is a method for monitoring efficacy of a vaccine, following administration thereof, by determining levels of antibodies in the subject which are specific for the vaccine itself, or a large molecule of which the vaccine is a part.

The identification of the subject proteins as being implicated in pathological conditions such as cancer also suggests a number of therapeutic approaches in addition to those discussed supra. The experiments set forth supra establish that antibodies are produced in response to expression of the protein. Hence, a further embodiment of the invention is the treatment of conditions which are characterized by aberrant or abnormal levels of one or more of the proteins, via administration of antibodies, such as humanized antibodies, antibody fragments, and so forth. These may be tagged or labelled with appropriate cytostatic or cytotoxic reagents.

T cells may also be administered. It is to be noted that the T cells may be elicited in vitro using immune responsive cells such as dendritic cells, lymphocytes, or any other immune responsive cells, and then reperfused into the subject being treated.

Note that the generation of T cells and/or antibodies can also be accomplished by administering cells, preferably treated to be rendered non-proliferative, which present relevant T cell or B cell epitopes for response, such as the epitopes discussed supra.

The therapeutic approaches may also include antisense therapies, wherein an antisense molecule, preferably from 10 to 100 nucleotides in length, is administered to the subject either "neat" or in a carrier, such as a liposome, to facilitate incorporation into a cell, followed by inhibition of expression of the protein. Such antisense sequences may also be incorporated into appropriate vaccines, such as in viral vectors (e.g., Vaccinia), bacterial constructs, such as variants of the known BCG vaccine, and so forth.

Also a part of the inventions are Peptides, such as those set forth in Figure 1, and those which have as a core sequence

PQSPLQI (SEQ ID NO.: 3)

These peptides may be used therapeutically, via administration to a patient who expresses CT7 in connection with a pathology, as well as diagnostically, i.e., to determine if relevant antibodies are present and so forth.

Other features and applications of the invention will be clear to the skilled artisan, and need not be set forth herein. The terms and expression which have been employed are used as terms of description and not of limitation, and there is no intention in the use of such terms and expression of excluding any equivalents of the features shown and described or portions thereof, it being recognized that various modifications are possible within the scope of the invention.

We claim:

1. Isolated nucleic acid molecule which encodes a cancer associated antigen, whose amino acid sequence is identical to the amino sequence encoded by nucleotides 287 to 3714 of SEQ ID NO: 1.
2. The isolated nucleic acid molecule of claim 1, consisting of nucleotides 287-3714 of SEQ ID NO: 1.
3. The isolated nucleic acid molecule of claim 1, consisting of anywhere from nucleotide 1 through nucleotide 4265 of SEQ ID NO: 1, with the proviso that said isolated nucleic acid molecule contains at least nucleotides 287-3714 of SEQ ID NO: 1.
4. Expression vector comprising the isolated nucleic acid molecule of claim 1, operably linked to a promoter.
5. Expression vector comprising the isolated nucleic acid molecule of claim 3, operably linked to a promoter.
6. Eukaryotic cell line or prokaryotic cell strain, transformed or transfected with the expression vector of claim 4.
7. Eukaryotic cell line or prokaryotic cell strain, transformed or transfected with the expression vector of claim 5.
8. Isolated cancer associated antigen comprising all or part of the amino acid sequence encoded by nucleotides 287-3714 of SEQ ID NO: 1.
9. Eukaryotic cell line or prokaryote cell strain, transformed or transfected with the isolated nucleic acid molecule of claim 1.
10. The eukaryotic cell line of claim 9, wherein said cell line is also transfected with a nucleic acid molecule coding for a cytokine.

11. The eukaryotic cell line of claim 10, wherein said cell line is further transfected by a nucleic acid molecule coding for an HLA molecule.
12. The eukaryotic cell line of claim 10, wherein said cytokine is an interleukin.
13. The biologically pure culture of claim 12, wherein said interleukin is IL-2, IL-4 or IL-12.
14. The eukaryotic cell line of claim 9, wherein said cell line has been rendered non-proliferative.
15. The eukaryotic cell line of claim 9, wherein said cell line is a fibroblast cell line.
16. Expression vector comprising a mutated or attenuated virus and the isolated nucleic acid molecule of claim 1.
17. The expression vector of claim 16, wherein said virus is adenovirus or vaccinia virus.
18. The expression vector of claim 17, wherein said virus is vaccinia virus.
19. The expression vector of claim 17, wherein said virus is adenovirus.
20. Expression system useful in transfecting a cell, comprising (i) a first vector containing a nucleic acid molecule which codes for the isolated cancer associated antigen of claim 8 and (ii) a second vector selected from the group consisting of (a) a vector containing a nucleic acid molecule which codes for an MHC or HLA molecule which presents an antigen derived from said cancer associated antigen and (b) a vector containing a nucleic acid molecule which codes for an interleukin.
21. Isolated cancer associated antigen comprising the amino acid sequence encoded by nucleotides 287-3714 of SEQ ID NO: 1.

22. Immunogenic composition comprising the isolated antigen of claim 21, and a pharmaceutically acceptable adjuvant.
23. The immunogenic composition of claim 22, wherein said adjuvant is a cytokine, a saponin, or GM-CSF.
24. Immunogenic composition comprising at least one peptide consisting of an amino acid sequence of from 8 to 12 amino acids concatenated to each other in the isolated cancer associated antigen of claim 21, and a pharmaceutically acceptable adjuvant.
25. The immunogenic composition of claim 24, wherein said adjuvant is a saponin, a cytokine, or GM-CSF.
26. The immunogenic composition of claim 24, wherein said composition comprises a plurality of peptides which complex with a specific MHC molecule.
27. Isolated peptide derived from the amino acid sequence encoded by SEQ ID NO: 1, wherein said isolated peptide binds to an HLA molecule, is a nonamer, decamer or undecamer, and comprises the amino acid sequence of SEQ ID NO: 3, from one to three additional N-terminal amino acid, and up to four additional C terminal amino acids.
28. Immunogenic composition which comprises at least one expression vector which encodes for a peptide derived from the amino acid sequence encoded by SEQ ID NO: 1, and an adjuvant or carrier.
29. The immunogenic composition of claim 28, wherein said at least one expression vector codes for a plurality of peptides.
30. Vaccine useful in treating a subject afflicted with a cancerous condition comprising the isolated cell line of claim 11 and a pharmacologically acceptable adjuvant.

31. The vaccine of claim 30, wherein said cell line has been rendered non-proliferative.
32. The vaccine of claim 31, wherein said cell line is a human cell line.
33. A composition of matter useful in treating a cancerous condition comprising a non-proliferative cell line having expressed on its surface a peptide derived from the amino acid sequence encoded by SEQ ID NO: 1.
34. The composition of matter of claim 33, wherein said cell line is a human cell line.
35. A composition of matter useful in treating a cancerous condition, comprising (i) a peptide derived from the amino acid sequence encoded by SEQ ID NO: 1, (ii) an MHC or HLA molecule, and (iii) a pharmaceutically acceptable carrier.
36. Isolated antibody which is specific for the antigen of claim 21.
37. The isolated antibody of claim 36, wherein said antibody is a monoclonal antibody.
38. Method for screening for cancer in a sample, comprising contacting said sample with a nucleic acid molecule which hybridizes to all or part of SEQ ID NO: 1, and determining hybridization as an indication of cancer cells in said sample.
39. A method for screening for cancer in a sample, comprising contacting said sample with the isolated antibody of claim 36, and determining binding of said antibody to a target as an indicator of cancer.
40. Method for diagnosing a cancerous condition in a subject, comprising contacting an immune reactive cell containing sample of said subject to a cell line transfected with the isolated nucleic acid molecule of claim 1, and determining interaction of said transfected cell line with said immunoreactive cell, said interaction being indicative of said cancer condition.

41. A method for determining regression, progression or onset of a cancerous condition comprising monitoring a sample from a patient with said cancerous condition for a parameter selected from the group consisting of (i) CT7 protein, (ii) a peptide derived from CT7 protein (iii) cytolytic T cells specific for said peptide and an MHC molecule with which it non-covalently complexes, and (iv) antibodies specific for said CT7 protein, wherein amount of said parameter is indicative of progression or regression or onset of said cancerous condition.
42. Method of claim 41, wherein said sample is a body fluid or exudate.
43. Method of claim 41, wherein said sample is a tissue.
44. Method of claim 41, comprising contacting said sample with an antibody which specifically binds with said protein or peptide.
45. Method of claim 44, wherein said antibody is labelled with a radioactive label or an enzyme.
46. Method of claim 44, wherein said antibody is a monoclonal antibody.
47. Method of claim 41, comprising amplifying RNA which codes for said protein.
48. Method of claim 47, wherein said amplifying comprises carrying out polymerase chain reaction.
49. Method of claim 41, comprising contacting said sample with a nucleic acid molecule which specifically hybridizes to a nucleic acid molecule which codes for or expresses said protein.
50. Method of claim 41, comprising assaying said sample for shed protein.
51. Method of claim 41, comprising assaying said sample for antibodies specific for said CT7 protein, by contacting said sample with CT7 protein.

52. Method for diagnosing a cancerous condition comprising assaying a sample taken from a subject for an immunoreactive cell specific for a peptide derived from CT7, complexed to an MHC molecule, presence of said immunoreactive cell being indicative of said cancerous condition.
53. An isolated nucleic acid molecule which encodes a protein and which has a complementary sequence which hybridizes, under stringent conditions, to at least one of the nucleotide sequences set forth at SEQ ID NO: 5, 6, 7 or 8.
54. The isolated nucleic acid molecule of claim 53, wherein said protein is the protein encoded by the nucleotide sequence of SEQ ID NO: 5, 6, 7 or 8.
55. The isolated nucleic acid molecule of claim 53, selected from the group consisting of nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 5, 6, 7 or 8.
56. Expression vector comprising the isolated nucleic acid molecule of claim 54, operably linked to a promoter.
57. Expression vector comprising the isolated nucleic acid molecule of claim 55, operably linked to a promoter.
58. Recombinant cell comprising the expression vector of claim 56.
59. Recombinant cell comprising the expression vector of claim 57.
60. Recombinant cell comprising the isolated nucleic acid molecule of claim 54.
61. Recombinant cell comprising the isolated nucleic acid molecule of claim 55.
62. Recombinant cell of claim 58, further comprising an expression vector which contains a nucleic acid molecule encoding a cytokine, operably linked to a promoter.
63. Recombinant cell of claim 59, further comprising an expression vector which contains a nucleic acid molecule encoding a cytokine, operably linked to a promoter.

64. Recombinant cell of claim 60, further comprising a nucleic acid molecule which encodes a cytokine.
65. Recombinant cell of claim 61, further comprising a nucleic acid molecule which encodes a cytokine.
66. The recombinant cell of claim 62, 63, 64, or 65, wherein said cytokine is interleukin.
67. The recombinant cell of claim 66, wherein said interleukin is 1L-2, 1L-4, or 1L-12.
68. The recombinant cell of claim 58, 59, 60, or 61, wherein said recombinant cell is a eukaryotic cell.
69. The recombinant cell of claim 68, which has been rendered non-proliferative.
70. The recombinant cell of claim 68, wherein said cell is a fibroblast.
71. Expression vector comprising a mutated or attenuated virus and the isolated nucleic acid molecule of claim 53, 54 or 55.
72. The expression vector of claim 71, wherein said virus is adenovirus, adeno associated virus, or vaccinia virus.
73. Expression system useful in making a recombinant cell, comprising:
 - (i) a first vector which encodes the protein encoded by the isolated nucleic acid molecule of claim 53, 54 or 55, and
 - (ii) a second vector which either (a) encodes an MHC or HLA molecule or (b) encodes an interleukin.
74. An isolated cancer associated antigen comprising the amino acid sequence encoded by SEQ ID NO: 5, 6, 7 or 8.

75. Composition comprising the isolated cancer associated antigen of claim 74, and a pharmaceutically acceptable adjuvant.
76. The composition of claim 75, wherein said adjuvant is a cytokine, a saponin, or GM-CSF.
77. Composition comprising at least one peptide consisting of an amino acid sequence of from 8 to 25 amino acids concatenated to each other in the isolated cancer associated antigen of claim 74, and a pharmaceutically acceptable adjuvant.
78. The composition of claim 77, wherein said adjuvant is a saponin, a cytokine, or GM-CSF.
79. The composition of claim 77, comprising a plurality of MHC binding peptides.
80. Composition comprising an expression vector which encodes at least one peptide consisting of an amino acid sequence of from 8 to 25 amino acids concatenated to each other in the isolated cancer associated antigen of claim 74, and pharmaceutically acceptable adjuvant.
81. The composition of claim 80, wherein said expression vector encodes a plurality of peptides.
82. Composition useful in treating a subject afflicted with a cancer, comprising the recombinant cell of claim 69 and a pharmacologically acceptable adjuvant.
83. The composition of claim 82, wherein said recombinant cell expresses an HLA or MHC molecule.
84. The composition of claim 82, wherein said recombinant cell is a human cell.
85. The composition of claim 77, further comprising at least one MHC or HLA molecule.

86. Isolated antibody which specifically binds to the isolated cancer associated antigen of claim 74.
87. The isolated antibody of claim 86, wherein said antibody is a monoclonal antibody.
88. A method for screening for possible presence of a pathological condition, comprising assaying a sample from a patient believed to have a pathological condition for antibodies specific to at least one of the cancer associated antigens encoded by SEQ ID NOS: 4, 5, 6, 7 or 8, presence of said antibodies being indicative of possible presence of said pathological condition.
89. The method of claim 88, wherein said pathological condition is cancer.
90. The method of claim 89, wherein said cancer is melanoma.
91. The method of claim 90, further comprising contacting said sample to purified cancer associated antigen encoded by SEQ ID NO: 4, 5, 6, 7 or 8.
92. A method for screening for possible presence of a pathological condition in a subject, comprising assaying a sample taken from said subject for expression of a nucleic acid molecule, the nucleotide sequence of which comprises SEQ ID NO: 5, 6, 7 or 8, expression of said nucleic acid molecule being indicative of possible presence of said pathological condition.
93. The method of claim 92, wherein said pathological condition is cancer.
94. The method of claim 92, comprising determining expression via polymerase chain reaction.
95. The method of claim 92, comprising determining expression by contacting said sample with at least one of SEQ ID NO: 11, 12, 13 or 14.

96. A method for determining regression, progression or onset of a cancerous condition comprising monitoring a sample from a patient with said cancerous condition for a parameter selected from the group consisting of (i) a cancer associated antigen encoded by SEQ ID NO: 3, 4, 5 or 6, (ii) a peptide derived from said cancer associated antigen, (iii) cytolytic T cells specific for said peptide and an MHC molecule with which it non-covalently complexes, and (iv) antibodies specific for said cancer associated antigen, wherein amount of said parameter is indicative of progression or regression or onset of said cancerous condition.

97. The method of claim 96, wherein said sample is a body fluid or exudate.

98. The method of claim 96, wherein said sample is a tissue.

99. The method of claim 96, comprising contacting said sample with an antibody which specifically binds with said protein or peptide.

100. The method of claim 99, wherein said antibody is labelled with a radioactive label or an enzyme.

101. The method of claim 99, wherein said antibody is a monoclonal antibody.

102. The method of claim 96, comprising amplifying RNA which codes for said protein.

103. The method of claim 102, wherein said amplifying comprises carrying out polymerase chain reaction.

104. The method of claim 96, comprising contacting said sample with a nucleic acid molecule which specifically hybridizes to a nucleic acid molecule which codes for or expresses said protein.

105. The method of claim 96, comprising assaying said sample for shed cancer associated antigen.

106. The method of claim 96, comprising assaying said sample for antibodies specific for said cancer associated antigen, by contacting said sample with said cancer associated antigen.

107. Method for screening for a cancerous condition comprising assaying a sample taken from a subject for an immunoreactive cell specific for a peptide derived from a cancer associated antigen encoded by SEQ ID NO: 4, 5, 6, 7 or 8 complexed to an MHC molecule, presence of said immunoreactive cell being indicative of said cancerous condition.

108. An isolated nucleic acid molecule consisting of a nucleotide sequence defined by SEQ ID NO: 9, 10, 11, 12, 13 or 14.

109. Kit useful in determining expression of a cancer associated antigen, comprising a separate portion of each of (i) the nucleotide sequences defined by SEQ ID NOS: 9 and 10, (ii) the nucleotide sequences defined by SEQ ID NOS: 11 and 12, and (iii) the nucleotide sequences defined by SEQ ID NOS: 13 and 14.

<110> Chen, Yao-Tseng
 Gure, Ali
 Tsang, Solam
 Stockert, Elisabeth
 Jager, Elke
 Knuth, Alexander
 Old, Lloyd J.

<120> Isolated Nucleic Acid Molecules Encoding Cancer Associated Antigen, The Antigens Per Se, And Uses Thereof

<130> LUD 5538.1 PCT

<140> PCT/US 99/05766

<141> 1998 - 04 - 17

<160> 3

<210> 1
 <211> 4265
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 1

GTCTGAAGGA	CCTGAGGCAT	TTTGTGACGA	GGATCGTCTC	AGGTCA	GGCG	AGGGAGGAGA	60
CTTATAGACC	TATCCAGTCT	TCAGGTTGCT	CCAGAAAGCA	GGAGTTGAAG	ACCTGGGTGT	120	
GAGGGACACA	TACATCCTAA	AAGCACCACA	GCAGAGGAGG	CCCAGGCAGT	GCCAGGAGTC	180	
AAGGTTCCCA	GAAGACAAAC	CCCCTAGGAA	GACAGGCGAC	CTGTGAGGCC	CTAGAGCACC	240	
ACCTTAAGAG	AAGAAGAGCT	CTAAGCCGGC	CTTTGTCA	GCCATCATGG	GGGACAAGGA	300	
TATGCCTACT	GCTGGGATGC	CGAGTCTTCT	CCAGAGTTCC	TCTGAGAGTC	CTCAGAGTTG	360	
TCCTGAGGGG	GAGGACTCCC	AGTCTCTCT	CCAGATCCC	CAGAGTCTC	CTGAGAGCGA	420	
CGACACCCCTG	TATCCTCTCC	AGAGTCTCA	GAGTCGTTCT	GAGGGGAGG	ACTCCTCGGA	480	
TCCCTCTCCAG	AGACCTCTCG	AGGGGAAGGA	CTCCCTAGCT	CCTCTCCAGA	TTCCCCAGAG	540	
TTCTCCTGAG	GGCGACGACA	CCCAGTCTCC	TCTCCAGAAT	TCTCAGAGTT	CTCCTGAGGG	600	
GAAGGACTCC	CTGTCTCTC	TAGAGATTTC	TCAGAGCCCT	CCTGAGGGTG	AGGATGTCCA	660	
GTCTCCTCTG	CAGAATCTCG	CGAGTCTCTT	CTTCTCCTCT	GCTTATTGA	GTATTTCCA	720	
GAGTTCCCTCT	GAGAGTATTC	AAAGTCTTT	TGAGGGTTTT	CCCCAGTCTG	TTCTCCAGAT	780	
TCCGTGAGC	GCCGCCTCTT	CCTCCACTTT	AGTGAGTATT	TTCCAGAGTT	CCCCTGAGAG	840	
TACTCAAAGT	CCTTTTGAGG	GTTTCCCCA	GTCTCCACTC	CAGATTCTG	TGAGCCGCTC	900	
CTTCTCCTCC	ACTTTATTGA	GTATTTCCA	GAGTCCCCCT	GAGAGAAC	AGAGAACCTC	960	
TGAGGGTTTT	GCACAGTCTC	CTCTCCAGAT	TCTGTGAGC	TCCTCCTCGT	CCTCCACTTT	1020	
ACTGAGTCTT	TTCCAGAGTT	CCCCTGAGAG	AACTCAGAGT	ACTTTTGAGG	GTTTCCCCA	1080	
GTCTCCACTC	CAGATTCTG	TGAGCCGCTC	CTTCTCCTCC	ACTTTATTGA	GTATTTCCA	1140	
GAGTTCCCTCT	GAGAGAAC	AGAGTACTTT	TGAGGGTTTT	CCCCAGTCTC	CTCTCCAGAT	1200	
TCCGTGAGC	CCCTCCTCT	CCTCCACTTT	AGTGAGTATT	TTCCAGAGTT	CCCCTGAGAG	1260	
AACTCAGAGT	ACTTTTGAGG	GTTTCCCCA	GTCTCCTCTC	CAGATTCTG	TGAGCTCCTC	1320	
CTTCTCCTCC	ACTTTATTGA	GTCTTTCCA	GAGTCCCCCT	GAGAGAAC	AGAGTACTTT	1380	
TGAGGGTTTT	CCCCAGTCTC	CTCTCCAGAT	TCTGAGGAGC	CCCTCCTCT	CCTCCACTTT	1440	
ACTGAGTCTT	TTCCAGAGTT	CCCCTGAGAG	AACTCAGAGT	ACTTTTGAGG	GTTTCCCCA	1500	
GTCTCCTCTC	CAGATTCTC	TGACCTCCTC	CTTCTCCTCT	ACTTTATTGA	GTATTTACA	1560	
GAGTTCTCTCT	GAGAGTCTC	AAAGTCTTT	TGAGGGTTTT	CCCCAGTCTC	CTCTCCAGAT	1620	
TCCGTGAGC	TCCTCTTCT	CCTACACTTT	ATTGAGTCTT	TTCCAGAGTT	CCCCTGAGAG	1680	
AACTCACAGT	ACTTTTGAGG	GTTTCCCCA	GTCTCCTCTC	CAGATTCTG	TGAGCTCCTC	1740	
CTCTCCTCTC	TCCACTTTAT	TGAGTCTTTT	CCAGAGTTCC	CCTGAGTGT	CTCAAAGTAC	1800	
TTTGAGGGT	TTTCCCCAGT	CTCCTCTCCA	GATTCTCAG	AGTCCTCCTG	AAGGGGAGAA	1860	
TACCCATTCT	CCTCTCCAGA	TTGTTCCAAG	TCTTCTGAG	TGGGAGGACT	CCCTGTCTCC	1920	
TCACTACTTT	CCTCAGAGCC	CTCCTCAGGG	GGAGGACTCC	CTATCTCCTC	ACTACTTCC	1980	
TCAGAGCCCT	CCTCAGGGGG	AGGACTCCCT	GTCTCCTCAC	TACTTCTC	AGAGCCCTCA	2040	
GGGGGAGGAC	TCCCTGTCTC	CTCACTACTT	TCTCAGAGC	CCTCCTCAGG	GGGAGGACTC	2100	
CATGTCTCT	CTCTACTTT	CTCAGAGTCC	TCTTCAGGGG	GAGGAATCC	AGTCTTCTCT	2160	
CCAGAGCCCT	GTGAGCATCT	GTCCTCCTC	CACTCCATCC	AGTCTTCCCC	AGAGTTTCCC	2220	
TGAGAGTTCT	CAGAGTCTC	CTGAGGGGCC	TGTCAGTCT	CCTCTCCATA	GTCTCAGAG	2280	

CCCTCCCTGAG	GGGATGCACT	CCCAATCTCC	TCTCCAGAGT	CCTGAGAGTG	CTCCCTGAGGG	2340
GGAGGATTCC	CTGCTCTCTC	TCCAAATTCC	TCAGAGTCCT	CTTGAGGGAG	AGGACTCCCT	2400
GTCTTCTCTC	CATTTCCCTC	AGAGTCCTCC	TGAGTGGGAG	GACTCCCTCT	CTCCTCTCCA	2460
CTTTCCTCAG	TTTCCTCTCTC	AGGGGGAGGA	CTTCCAGTCT	TCTCTCCAGA	GTCCTGTGAG	2520
TATCTGCTCC	TCCTCCACTT	CTTIGAGTCT	TCCCCAGAGT	TTCCCTGAGA	GTCCTCAGAG	2580
TCCTCCTGAG	GGGCCTGCTC	AGTCTCTCTC	CCAGAGACCT	GTCAGCTCCT	TCTTCTCTTA	2640
CACTTTAGCG	AGTCTTCTCC	AAAGTCCCCA	TGAGAGTCCT	CAGAGTCCTC	CTGAGGGGCC	2700
TGCCAGTCT	CCTCTCCAGA	GTCCTGTGAG	CTCCTTCCCC	TCCTCCACTT	CATCGAGTCT	2760
TTCCCAGAGT	TCTCTGTGA	GTCCTTCCCC	CTCCTCCACT	TCATCGAGTC	TTTCCAAGAG	2820
TTCCCCTGAG	AGTCTCTCTC	AGAGTCCTGT	GATCTCCTTC	TCCTCCTCCA	CTTCATTGAG	2880
CCCATTCACT	GAAGAGTCCA	GCAGGCCAGT	AGATGAATAT	ACAAGTTCCT	CAGACACCTT	2940
GCTAGAGAGT	GATTCTTGTGA	CAGACAGCGA	GTCCTTGATA	GAGAGCGAGC	CCTTGTTCAC	3000
TTATACACTG	GATGAAAAGG	TGGAGGAGT	GGCGCGGTTT	CTTCTCCTCA	AATATCAAGT	3060
GAAGCAGCCT	ATCACAAAGG	CAGAGATGCT	GACGAATGTC	ATCAGCAGGT	ACACGGGTC	3120
CTTTCCTGTG	ATCTCTAGGA	AAGCCCGTGA	GTTCATAGAG	ATATTTTTG	GCATTTCCCT	3180
GAGAGAAAGT	GACCTGTATG	ACTCTTATGT	CTTGTAAAC	ACATTAGACC	TCACCTCTGA	3240
GGGGTGTCTG	AGTGATGAGC	AGGGCATGTC	CCAGAACCGC	CTCTGTATTG	TTATTCTGAG	3300
TATCATCTTC	ATAAAAGGGCA	CCTATGCCTC	TGAGGAGGTC	ATCTGGGATG	TGCTGAGTGG	3360
AATAGGGGTG	CGTGTGGGA	GGGAGCACTT	TGCTTTGGG	GAGCCCAGGG	AGCTCCTCAC	3420
TAAGTTTGG	GTGCAGGAAC	ATTACCTAGA	GTACCGGGAG	GTGCCCAACT	CTTCTCTCC	3480
TCGTTACGAA	TTCTGTGGG	GTCCAAGAGC	TCATTAGAA	GTCATTAAGA	GGAAAGTAGT	3540
AGAGTTTTTG	GCCATGCTAA	AGAATACCGT	CCCTATTACC	TTTCCATCCT	CTTACAAGGA	3600
TGCTTTGAAA	GATGTGGAAG	AGAGAGCCCC	GGCCATAATT	GACACCACAG	ATGATTGAC	3660
TGCCACAGAA	AGTGAAGCT	CCAGTGTAT	GTCCCCCAGC	TTCTCTTCTG	AGTGAAGTCT	3720
AGGGCAGATT	CTTCCCTCTG	AGTTTGAAGG	GGGCAGTCGA	TTTCTACGTT	GGTGGAGGGC	3780
CTGGTTGAGG	CTGGAGAGAA	CACAGTGCTA	TTTGCATTTC	TGTCCCATAT	GGGTAGTTAT	3840
GGGGTTTAC	TGTTTTACTT	TTGGGTATT	TTCAAATGCT	TTTCTTATTA	ATAACAGGTT	3900
TAATAGCTT	CAGAATCTCA	GTATTATGCAC	ATGAGTCGCA	CATGTATTGC	TGTTTTCTG	3960
GTAAAGAGT	AACAGTTGA	TATTTTGTAA	AAACAAAAAC	ACACCCAAAC	ACACACATT	4020
GGGAAACCT	TCTGCTCTCAT	TTTGTGATGT	GTCACAGGTT	AATGTGGTGT	TACTGTAGGA	4080
ATTTCTTGA	AACTGTGAAG	GAACCTGCA	GTAAATAGT	GGAATAAGT	AAAGGATTGT	4140
TAATGTTG	ATTTCTCAG	GTCCCTTGTAG	CTGTTGTCT	TGAAAACCTAA	AGATACATAC	4200
CTGGTTTGCT	TGGCTTACGT	AAGAAAGTCG	AAGAAAGTAA	ACTGTAATAA	ATAAAAGTGT	4260
CACTG						4265

<210> 2
<211> 1142
<212> PRT
<213> Homo sapiens
<220>
<400> 2

Met	Gly	Asp	Lys	Asp	Met	Pro	Thr	Ala	Gly	Met	Pro	Ser	Leu	Leu	Gln
					5				10					15	
Ser	Ser	Ser	Glu	Ser	Pro	Gln	Ser	Cys	Pro	Glu	Gly	Glu	Asp	Ser	Gln
					20				25					30	
Ser	Pro	Leu	Gln	Ile	Pro	Gln	Ser	Ser	Pro	Glu	Ser	Asp	Asp	Thr	Leu
							35			40				45	
Tyr	Pro	Leu	Gln	Ser	Pro	Gln	Ser	Arg	Ser	Glu	Gly	Glu	Asp	Ser	Ser
							50		55		60				
Asp	Pro	Leu	Gln	Arg	Pro	Pro	Glu	Gly	Lys	Asp	Ser	Gln	Ser	Pro	Leu
							65		70		75				80
Gln	Ile	Pro	Gln	Ser	Ser	Pro	Glu	Gly	Asp	Asp	Thr	Gln	Ser	Pro	Leu
							85			90				95	

Gln	Asn	Ser	Gln	Ser	Ser	Pro	Glu	Gly	Lys	Asp	Ser	Leu	Ser	Pro	Leu
100							105					110			
Glu	Ile	Ser	Gln	Ser	Pro	Pro	Glu	Gly	Glu	Asp	Val	Gln	Ser	Pro	Leu
115							120					125			
Gln	Asn	Pro	Ala	Ser	Ser	Phe	Phe	Ser	Ser	Ala	Leu	Leu	Ser	Ile	Phe
130						135					140				
Gln	Ser	Ser	Pro	Glu	Ser	Ile	Gln	Ser	Pro	Phe	Glu	Gly	Phe	Pro	Gln
145						150				155					160
Ser	Val	Leu	Gln	Il	Pro	Val	Ser	Ala	Ala	Ser	Ser	Ser	Thr	Leu	Val
					165				170					175	

SUBSTITUTE SHEET (Rule 26)

Ser Ile Phe Gln Ser Ser Pro Glu Ser Thr Gln Ser Pro Phe Glu Gly
 180 185 190
 Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser Arg Ser Phe Ser Ser
 195 200 205
 Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro Glu Arg Ser Gln Arg Thr
 210 215 220
 Ser Glu Gly Phe Ala Gln Ser Pro Leu Gln Ile Pro Val Ser Ser Ser
 225 230 235 240
 Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr
 245 250 255
 Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val
 260 265 270
 Ser Arg Ser Phe Ser Ser Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro
 275 280 285
 Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Ala Gln Ser Pro Leu Gln
 290 295 300
 Ile Pro Val Ser Pro Ser Phe Ser Ser Thr Leu Val Ser Ile Phe Gln
 305 310 315 320
 Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser
 325 330 335
 Pro Leu Gln Ile Pro Val Ser Ser Ser Phe Ser Ser Thr Leu Leu Ser
 340 345 350
 Leu Phe Gln Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe
 355 360 365
 Pro Gln Ser Pro Leu Gln Ile Pro Gly Ser Pro Ser Phe Ser Ser Thr
 370 375 380
 Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr His Ser Thr Phe
 385 390 395 400
 Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Met Thr Ser Ser Phe
 405 410 415
 Ser Ser Thr Leu Leu Ser Ile Leu Gln Ser Ser Pro Glu Ser Ala Gln
 420 425 430
 Ser Ala Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser
 435 440 445
 Ser Ser Phe Ser Tyr Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu
 450 455 460
 Arg Thr His Ser Thr Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile
 465 470 475 480
 Pro Val Ser Ser Ser Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln
 485 490 495
 Ser Ser Pro Glu Cys Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser
 500 505 510
 Pro Leu Gln Ile Pro Gln Ser Pro Pro Glu Gly Glu Asn Thr His Ser
 515 520 525
 Pro Leu Gln Ile Val Pro Ser Leu Pro Glu Trp Glu Asp Ser Leu Ser
 530 535 540
 Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Leu Ser
 545 550 555 560
 Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Leu Ser
 565 570 575
 Pro His Tyr Phe Pro Gln Ser Pro Gln Gly Glu Asp Ser Leu Ser Pro
 580 585 590
 His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Met Ser Pro
 595 600 605
 Leu Tyr Phe Pro Gln Ser Pro Leu Gln Gly Glu Glu Phe Gln Ser Ser
 610 615 620
 Leu Gln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Pro Ser Ser Leu
 625 630 635 640
 Pro Gln Ser Phe Pro Glu Ser Ser Gln Ser Pro Pro Glu Gly Pro Val
 645 650 655
 Gln Ser Pro Leu His Ser Pro Gln Ser Pro Pro Glu Gly Met His Ser
 660 665 670
 Gln Ser Pro Leu Gln Ser Pro Glu Ser Ala Pro Glu Gly Glu Asp Ser
 675 680 685

Leu Ser Pro Leu Gln Ile Pro Gln Ser Pro Leu Glu Gly Glu Asp Ser
 690 695 700
 Leu Ser Ser Leu His Phe Pro Gln Ser Pro Pro Glu Trp Glu Asp Ser
 705 710 715 720
 Leu Ser Pro Leu His Phe Pro Gln Phe Pro Pro Gln Gly Glu Asp Phe
 725 730 735
 Gln Ser Ser Leu Gln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Ser
 740 745 750
 Leu Ser Leu Pro Gln Ser Phe Pro Glu Ser Pro Gln Ser Pro Pro Glu
 755 760 765
 Gly Pro Ala Gln Ser Pro Leu Gln Arg Pro Val Ser Ser Phe Phe Ser
 770 775 780
 Tyr Thr Leu Ala Ser Leu Leu Gln Ser Ser His Glu Ser Pro Gln Ser
 785 790 795 800
 Pro Pro Glu Gly Pro Ala Gln Ser Pro Leu Gln Ser Pro Val Ser Ser
 805 810 815
 Phe Pro Ser Ser Thr Ser Ser Leu Ser Gln Ser Ser Pro Val Ser
 820 825 830
 Ser Phe Pro Ser Ser Thr Ser Ser Leu Ser Lys Ser Ser Pro Glu
 835 840 845
 Ser Pro Leu Gln Ser Pro Val Ile Ser Phe Ser Ser Ser Thr Ser Leu
 850 855 860
 Ser Pro Phe Ser Glu Glu Ser Ser Ser Pro Val Asp Glu Tyr Thr Ser
 865 870 875 880
 Ser Ser Asp Thr Leu Leu Glu Ser Asp Ser Leu Thr Asp Ser Glu Ser
 885 890 895
 Leu Ile Glu Ser Glu Pro Leu Phe Thr Tyr Thr Leu Asp Glu Lys Val
 900 905 910
 Asp Glu Leu Ala Arg Phe Leu Leu Lys Tyr Gln Val Lys Gln Pro
 915 920 925
 Ile Thr Lys Ala Glu Met Leu Thr Asn Val Ile Ser Arg Tyr Thr Gly
 930 935 940
 Tyr Phe Pro Val Ile Phe Arg Lys Ala Arg Glu Phe Ile Glu Ile Leu
 945 950 955 960
 Phe Gly Ile Ser Leu Arg Glu Val Asp Pro Asp Asp Ser Tyr Val Phe
 965 970 975
 Val Asn Thr Leu Asp Leu Thr Ser Glu Gly Cys Leu Ser Asp Glu Gln
 980 985 990
 Gly Met Ser Gln Asn Arg Leu Leu Ile Leu Ser Ile Ile Phe
 995 1000 1005
 Ile Lys Gly Thr Tyr Ala Ser Glu Glu Val Ile Trp Asp Val Leu Ser
 1010 1015 1020
 Gly Ile Gly Val Arg Ala Gly Arg Glu His Phe Ala Phe Gly Glu Pro
 1025 1030 1035 1040
 Arg Glu Leu Leu Thr Lys Val Trp Val Gln Glu His Tyr Leu Glu Tyr
 1045 1050 1055
 Arg Glu Val Pro Asn Ser Ser Pro Pro Arg Tyr Glu Phe Leu Trp Gly
 1060 1065 1070
 Pro Arg Ala His Ser Glu Val Ile Lys Arg Lys Val Val Glu Phe Leu
 1075 1080 1085
 Ala Met Leu Lys Asn Thr Val Pro Ile Thr Phe Pro Ser Ser Tyr Lys
 1090 1095 1100
 Asp Ala Leu Lys Asp Val Glu Glu Arg Ala Gln Ala Ile Ile Asp Thr
 1105 1110 1115 1120
 Thr Asp Asp Ser Thr Ala Thr Glu Ser Ala Ser Ser Ser Val Met Ser
 1125 1130 1135
 Pro Ser Phe Ser Ser Glu
 1140

<210> 3
 <211> 7
 <212> PRT
 <213> Homo sapiens
 <220>
 <400> 3

Pro Gln Ser Pro Leu Gln Ile
 1 5

<210> 4
 <211> 4159
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 4

GGTGGATGCG TTTGGGTTGT AGCTAGGCTT TTTCTTTCT TTCTCTTTA AAACACATCT 60
 AGACAAGGAA AAAACAAGCC TCGGATCTGA TTTTCACTC CTCGTTCTG TGCTTGGTTC 120
 TTACTGTGTT TGTGTATTTT AAAGGCAGA AGACGAGGGG AACAAAACCA GCTGGATCCA 180
 TCCATCACCG TGGGTGGTTT TAATTTTCG TTTTTCTCG TTATTTTTT TTAAACAACC 240
 ACTCTTCACA ATGAACAAAC TGTATATCGG AACACCTCAGC GAGAACGCCG CCCCCTCGGA 300
 CCTAGAAAGT ATCTTCAAGG ACGCCAAGAT CCCGGTGTGCG GGACCCTTCC TGGTGAAGAC 360
 TGGCTACGCG TTCGTGGACT GCCCCGACGA GAGCTGGCC CTCAAGGCCA TCGAGGCGCT 420
 TTCAGGTAAA ATAGAACTGC ACGGGAAACC CATAGAAGTT GAGCACTCGG TCCCCAAAAG 480
 GCAAAGGATT CGGAAAACCTC AGATAAGAAA TATCCCGCCT CATTACAGT GGGAGGTGCT 540
 GGATAGTTA CTAGTCCAGT ATGGAGTGGT GGAGAGCTGT GAGCAAGTGA ACAC TGACTC 600
 GGAAACTGCA GTTGTAAATG TAACCTATT CAGTAAGGAC CAAGCTAGAC AAGCACTAGA 660
 CAAACTGAAT GGATTTCACT TAGAGAATT CACCTTGAAAT GTAGCCTATA TCCCTGTGTA 720
 AATGGCCGCC CAGCAAAACC CTTGAGCA GCCCCGAGGT CGCCGGGGC TTGGGCAGAG 780
 GGGCTCCCTA AAGCAGGGGT CTCCAGGATC CGTATCCAAG CAGAAACCAT GTGATTGCC 840
 TCTGCGCTG CTGGTTCCCA CCCAATTGT TGAGGCCATC ATAGGAAAAG AAGGTGCCAC 900
 CATTGGAAC ATCACCAAAAC AGACCCAGTC TAAAATCGAT GTCCACCGTA AAGAAAATGC 960
 GGGGGCTGCT GAGAAGTCGA TTACTATCCT CTCTACTCCT GAAGGCACCT CTGGGCTTG 1020
 TAAGTCTATT CTGGAGATT TGCAAAAGGAGT ATAAAATCA CAGAAGAGAT 1080
 CCCCTTGAAG ATTTTAGCTC ATAATAACTT TGTTGGACGT CTTATTGGTA AAGAAGGAAG 1140
 AAATCTAAA AAAATTGAGC AAGACACAGA CACTAAAATC ACGATATCTC CATTGCAAGGA 1200
 ATTGACGCTG TATAATCCAG AACGCACAT TACAGTTAAA GGCAATGTT AGACATGTGC 1260
 CAAAGCTGAG GAGGAGATCA TGAAGAAAAT CAGGGAGTCT TATGAAAATG ATATTGCTTC 1320
 TATGAATCTT CAAGCACATT TAATTCTGG ATTAATCTG AACGCCTTGG GTCTGTTCCC 1380
 ACCCACTTCA GGGATGCCAC CTCCCACCTC AGGGCCCCCT TCAGCCATGA CTCCCTCC 1440
 CCCGCAGTTT GAGCAATCAG AAACGGAGAC TGTTCATCAG TTTATCCAG CTCTATCAGT 1500
 CGGTGCCATC ATCGGCAAGC AGGGCCAGCA CATCAAGCAG CTTTCTCGCT TTGCTGGAGC 1560
 TTCAATTAAAG ATTGCTCCAG CGGAAGCACC AGATGCTAAA TGAGGATGG TGATTATCAC 1620
 TGGACCACCA GAGGCTCAGT TCAAGGCTCA GGGAAAGATT TATGGAAAAA TTAAAGAAGA 1680
 AAACATTGTT AGCCTAAAG AAGAGGTGAA ACTTGAAGCT CATATCAGAG TGCCATCCTT 1740
 TGCTGCTGGC AGAGTTATTG GAAAAGGAGG CAAAACGGGT AATGAACCT AGAATTGTC 1800
 AAGTGCAGAA GTTGTGTCC CTCGTGACCA GACACCTGAT GAGAATGACC AAGTGGTTGT 1860
 CAAAATAACT GGTCACTTCT ATGCTTGCCA GGTGCCCCAG AGAAAAATTC AGGAAATTCT 1920
 GACTCAGGTA AACGAGCACC AACACAGAA GGCTCTGCAA AGTGGACAC CTCAGTCAAG 1980
 ACAGGAAGTAA AGGCTCAGGA AACAGCCCCAC CACAGAGGCA GATGCCAAC CAAAGACAGA 2040
 TTGCTTAACC AACAGATGGG CGCTGACCCCC CTATCCAGAA TCACATGCAC AAGTTTTAC 2100
 CTAGCCAGTT GTTCTGAGG ACCAGGCAAC TTTGAACCTC CTGTCTCTGT GAGAATGTAT 2160
 ACTTTATGCT CTCTGAAAATG TATGACACCC AGCTTTAAAA CAAACAAACA AACAAACAAA 2220
 AAAAGGGTGG GGGAGGGAGG GAAAGAGAAG AGCTCTGCAC TTCCCTTGT TGTAGTCTCA 2280
 CAGTATAACA GATATTCTAA TTCTCTTAA TATTCCCCCA TAATGCCAGA AATTGGCTTA 2340
 ATGATGCTT CACTAAATTG ATCAAATAGA TTGCTCTTAA ATCCAATTGT TAAAATTGGA 2400
 TCAGAATAAT TATCACAGGA ACTTAAATGT TAAGCCATTA GCATAGAAAA ACTGTTCTCA 2460
 GTTTTATTTT TACCTAACAC TAACATGAGT AACCTAAGGG AAGTGTGAA TGGTGTGCG 2520
 AGGGGTATTA AACGTGCATT TTACTCAAC TACCTCAGGT ATTCAAGTAAT ACAATGAAAA 2580
 GCAAAATTGT TCTTTTTTT TGAAAATTG ATATACTTAA TAATGATAGA AGTCCAACCG 2640
 TTTTTAAAAA ATAAATTTA AAATTAACA GCAATCAGCT AACAGGCAAA TTAAGATTTT 2700

TACTTCTGGC	TGGTGACAGT	AAAGCTGAA	AATTAATTTC	AGGGTTTTT	GAGGCCTTTG	2760
ACACAGTTAT	TAGTTAAATC	AAATGTTCAA	AAATACGGAG	CAGTGCCTAG	TATCTGGAGA	2820
GCAGCACTAC	CATTTATTCT	TTCATTTATA	GTGCGGAAAG	TTTTTGACGG	TACTAACAAA	2880
GTGGTCGAG	GAGATTTGG	AAACGGCTGGT	TTAAATGGCT	TCAGGAGACT	TCAGTTTTT	2940
GTTAGCTAC	ATGATTGAAT	GCATAATAAA	TGCTTTGTG	TTCTGACTAT	CAATACCTAA	3000
AGAAAGTGA	TCAGTGAAGA	GATGCAAGAC	TTTCAACTGA	CTGGCAAAAA	GCAAGCTTTA	3060
GCTTGTCTTA	TAGGATGCTT	AGTTGCCAC	TACACTTCAG	ACCAATGGGA	CAGTCATAGA	3120
TGGTGTGACA	GTGTTAAC	GCAACAAAAG	GCTACATTTC	CATGGGCCA	GCACTGTCT	3180
GAGCCTCACT	AAGCTATTTC	GAAGATTTT	AAAGCACTGAT	AAATTAACAA	AAAAAAAGAAA	3240
AAATTAGACT	CCACCTTAAG	TAGTAAAGTA	TAACAGGATT	TCTGTATACT	GTGCAATCAG	3300
TTCTTGAAA	AAAAAGTCAA	AAGATAGAGA	ATACAAGAAA	AGTTTTNGGG	ATATAATTG	3360
AATGACTGTG	AAAACATATG	ACCTTGATA	ACGAACACTCAT	TTGCTCACTC	CTTGACAGCA	3420
AAGCCCAGTA	CGTACAATTG	TGTTGGGTGT	GGGTGGTCTC	CAAGGCCACG	CTGCTCTCTG	3480
AATTGATTTT	TTGAGTTTG	GNTTGNAGA	TGATCACAGN	CATGTTACAC	TGATCTTNA	3540
GGACATATNT	TATAACCCCT	TAACACACAC	ATCCCTGCC	TCATTCTTAT	TTCGAGATGA	3600
ATTTCGATAC	AGACTAGATG	TCTTCTGAA	GATCAATTAG	ACATTNTGAA	AATGATTTAA	3660
AGTGTTTCC	TTAATGTTCT	CTGAAACAA	GTTCCTTTG	TAGTTTTAAC	AAAAAAAGTG	3720
CCCTTTTGT	CACTGGTTTC	TCCTAGCATT	CATGATTTT	TTTTCACACA	ATGAATTAAA	3780
ATTGCTAAAA	TCATGGACTG	GCTTCTGGT	TGGATTTCA	GTAAGATGTG	TTTAAGGCCA	3840
GAGCTTTCT	CAGTATTGAA	TTTTTTCCC	CAATATTGAA	TTTTTTAACAA	ATATACACAT	3900
AGGAGCTGCA	TTTAAACACT	GCTGGTTAA	ATTCTGTCAN	ATTTCACTTC	TAGCCTTTA	3960
GTATGGCNA	TCANAATTAA	CTTTTACTTA	AGCATTGTA	ATTTGGAGTA	TCTGGTACTA	4020
GCTAAGAAAT	AATTCNATAA	TTGAGTTTG	TACTCNCCAA	ANATGGTC	TTCCCTCATGN	4080
ATAATGTNCC	CCCAATGCAG	CTTCATTTTC	CAGANACCTT	GACGCAGGAT	AAATTTTTTC	4140
ATCATTAGG	TCCCCAAAAA					4159

<210> 5

<211> 1708

<212> DNA

<213> Homo sapiens

<220>

<400> 5

AGGGACGCTG	CCGCACCGCC	CCAGTTTACC	CCGGGGAGCC	ATCATGAAGC	TGAATGGCCA	60
CCAGTTGGAG	AACCATGCC	TGAAGGTCTC	CTACATCCCC	GATGAGCAGA	TAGCACAGGG	120
ACCTGAGAAAT	GGGCGCCGAG	GGGGCTTTGG	CTCTCGGGGT	CAGCCCCGCC	AGGGCTCACC	180
TGTGGCAGCG	GGGGCCCCAG	CCAAGCAGCA	GCAAGTGGAC	ATCCCCCTTC	GGCTCCTGGT	240
GCCCCACCCAG	TATGTTGGT	CCATTATTGG	CAAGGAGGGG	GCCACCATCC	GCAACATCAC	300
AAAACAGACC	CAGTCCAAGA	TAGACGTCA	TAGGAAGGAG	AACGCAAGGTG	CAGCTGAAAA	360
AGCCATCAGT	GTGCACTCCA	CCCCCTGAGGG	GTGCTCTTC	GCTTGTAAGA	TGATCTTGA	420
GATTATGCAT	AAAGAGGCTA	AGGACACAA	AAACGGCTGAC	GAGGTTCCCC	TGAAGATCCT	480
GGCCCATAAT	AACTTTGTAG	GGCGTCTCAT	TGGCAAGGAA	GGACGGAACC	TGAAGAAGGT	540
AGAGCAAGAT	ACCGAGACAA	AAATCACCAT	CTCCTCGTTG	CAAGACCTTA	CCCTTTACAA	600
CCCTGAGAGG	ACCATCACTG	TGAAGGGGGC	CATCGAGAAT	TGTTGCAGGG	CCGAGCAGGA	660
AATAATGAAG	AAAGTTCGGG	AGGCCTATGA	GAATGATGTG	GCTGCCATGA	GCTCTCACCT	720
GATCCCTGGC	CTGAACCTGG	CTGCTGTAGG	TCTTTTCCA	GCTTCATCCA	GCGCAGTCCC	780
GCCGCCTCCC	AGCAGCGTTA	CTGGGGCTGC	TCCCTATAGC	TCCCTTATGC	AGGCTCCCGA	840
GCAGGAGATG	GTGCAGGTGT	TTATCCCCGC	CCAGGCACTG	GGCGCCATCA	TCGGCAAGAA	900
GGGGCAGCAC	ATCAAACAGC	TCTCCGGTT	TGCCAGCGCC	TCCATCAAGA	TTGCACCACC	960
CGAAACACCT	GACTCCAAAG	TTCGTATGGT	TATCATCACT	GGACCGCCAG	AGGGCCAATT	1020
CAAGGCTCAG	GGAAAGAATCT	ATGGCAAAC	CAAGGAGGAG	AACTTCTTG	GTCCCAAGGA	1080
GGAAGTGAAG	CTGGAGACCC	ACATACTGT	GCCAGCATCA	GCAGCTGGCC	GGGTCAATTGG	1140
CAAAGGTGGA	AAAACGGTGA	ACGAGGTGCA	GAATTGACG	GCAGCTGAGG	TGGTAGTACC	1200
AAGAGACCAG	ACCCCTGATG	AGAACAGCCA	GGTCATCGT	AAAATCATCG	GACATTCTCA	1260
TGCCAGTCAG	ATGGCTAAC	GGAAAGATCCG	AGACATCTG	GCCCAGGTTA	AGCAGCAGCA	1320
TCAGAAGGG	CAGAGTAACC	AGGCCCAAGG	ACGGAGGAAG	TGACCAGCCC	CTCCCTGTCC	1380
CTTNGAGTCC	AGGACAACAA	CGGGCAGAAA	TGCGAGGTGT	GCTCTCCCCG	GCAGGCCTGA	1440
GAATGAGTGG	GAATCCGGGA	CACNTGGCC	GGGCTGTAGA	TCAGGTTTG	CCACTTGATT	1500
GAGAAAGATG	TTCCAGTGAG	GAACCTGTAT	CTNTCAGCCC	CAAACACCCA	CCCAATTGGC	1560
CCAACACTGT	NTGCCCTCTG	GGGTGTCAGA	AATTNTAGCG	CAAGGCACCTT	TTAACACGTGG	1620
ATTGTTAAA	GAAGCTCTCC	AGGCCCAACC	AAAGAGGGTGG	ATCACACCTC	AGTGGGAAGA	1680
AAAATAAAAT	TTCCTTCAGG	TTTTAAAAA				1708

<210> 6
 <211> 3412
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 6

GGCAGCGGAG	GAGGCGAGGA	GCGCCGGTA	CCGGGCCGGG	GGAGCCGCGG	GCTCTCGGGG	60
AAGAGACGGA	TGATGAACAA	GCTTACATC	GGGAACCTGA	GCCCCGCCGT	CACCGCCGAC	120
GACCTCCGGC	AGCTCTTGG	GGACAGGAAG	CTGCCCCCTGG	CGGGACAGGT	CCTGCTGAAG	180
TCCGGCTACG	CCTTCGTGGA	CTACCCCGAC	CAGAACTGGG	CCATCCGCGC	CATCGAGACC	240
CTCTCGGGTA	AAGTGGAAATT	GCATGGGAAA	ATCATGGAAG	TTGATTACTC	AGTCTCTAAA	300
AAGCTAAGGA	GCAGGAAAAT	TCAGATTGCA	AACATCCCTC	CTCACCTGCA	GTGGGAGGTG	360
TTGGATGGAC	TTTGGCTCA	ATATGGACA	GTGGAGAATG	TGGAACAAGT	CAACACAGAC	420
ACAGAAACCG	CCGTTGTCAA	CGTCACATAT	GCAACAAGAG	AAGAAGCAAA	AATAGCCATG	480
GAGAAGCTAA	GCAGGGCATCA	GTGGAGAAC	TACTCCTTCA	AGATTTCTTA	CATCCGGAT	540
GAAGAGGTGA	GCTCCCCTTC	GCCCCCTCG	CGAGGCCAGC	GTGGGGACCA	CTCTTCCCGG	600
GAGCAAGGCC	ACGCCCCCTGG	GGGCACTTCT	CAGGCCAGAC	AGATTGATT	CCCGCTGCGG	660
ATCTCGTCC	CCACCCAGTT	TGTTGGTGGC	ATCATCGGAA	AGGAGGGCT	GACCATAAAAG	720
AAACATCACTA	AGCAGACCCA	GTCCCCGGTA	GATATCCATA	GAAAAGAGAA	CTCTGGAGCT	780
GCAGAGAACG	CTGTCACCAT	CCATGCCACC	CCAGAGGGGA	CTTCTGAAGC	ATGCCGATG	840
ATTCTTGAAA	TCATGCAGAA	AGAGGCAGAT	GAGACCAAAC	TAGCGAAGA	GATTCCCTTG	900
AAAATCTTGG	CACACAATGG	CTTGGTTGGA	AGACTGATTG	GAAAAGAAGG	CAGAAATTG	960
AAGAAAATTG	AAACATGAAAC	AGGGACCAAG	ATAACAAATCT	CATCTTGC	GGATTGAGC	1020
ATATACAACC	CGGAAAAGAAC	CATCACTGTG	AAAGGGCACAG	TTGAGGCCTG	TGCCAGTGCT	1080
GAGATAGAGA	TTATGAAGAA	GCTGCGTGG	GCCTTGAAGA	ATGATATGCT	GGCTGTTAAC	1140
CAACAAGCCA	ATCTGATCCC	AGGGTGAAC	CTCAGCGCAC	TTGGCATTCT	TTCAACAGGA	1200
CTGTCGTC	TATCTCCACC	AGCAGGGCCC	CGGGAGGCTC	CCCCCGCTGC	CCCCTACAC	1260
CCCTTCACTA	CCCACCTCGG	ATACTCTCC	AGCCTGTACC	CCCATCACCA	GTGGGCCCG	1320
TTCCCGCATC	ATCACTCTTA	TCCAGAGCG	GAGATTGTGA	ATCTCTTCAT	CCCAACCCAG	1380
GCTGTGGCG	CCATCATCGG	GAAGAAGGGG	GCACACATCA	AAACAGCTGC	GAGATTGCC	1440
GGAGCCTCTA	TCAAGATTG	CCCTGCGGA	GGCCCAGACG	TCAGCGAAAG	GATGGTCATC	1500
ATCACCGGGC	CACCGGAAGC	CCAGTCAAG	GCCCAGGGAC	GGATCTTGG	GAAACTGAAA	1560
GAGGAAACT	TCTTAAACCC	CAAAGAAGAA	GTGAAGCTGG	AAGCGCATAT	CAGAGTGC	1620
TCTTCCACAG	CTGGCGGGGT	GATTGGCAA	GGTGGCAAGA	CCGTGAACGA	ACTGCAGAAC	1680
TTAACCGATG	CAGAAGTCAT	CGTGCCTCGT	GACCAACACG	CAGATGAAA	TGAGGAAGTG	1740
ATCGTCAGAA	TTATGGGCA	CTTCTTGT	AGCCAGACTG	CACAGCGAA	GATCAGGGAA	1800
ATTGTACAAAC	AGGTGAAGCA	GCAGGAGCG	AAATACCCCTC	AGGGAGTCG	CTCACAGCGC	1860
AGCAAGTGTG	GCTCCCACAG	GCACCAAGCAA	AAACACGGAT	GAATGTAGCC	CTTCCAACAC	1920
CTGACAGAAAT	GAGACCAAAC	GCAGCCAGCC	AGATCGGGAG	CAAACAAAG	ACCATCTGAG	1980
GAATGAGAAG	TCTGCGGAGG	CGGCCAGGG	CTCTGCCAG	GCCCTGAGAA	CCCCAGGGC	2040
CGAGGAGGGG	CGGGGAAGGT	CAGCCAGGT	TGCCAGAAC	ACCGAGCCCC	GCCTCCCGCC	2100
CCCCAGGGCT	TCTGCAGGCT	TCAGCCATCC	ACTTCACCAT	CCACTCGGAT	CTCTCCTGAA	2160
CTCCCACGAC	GCTATCCCTT	TTAGTTGAAC	TAACATAGGT	GAACGTGTT	AAAGCCAAGC	2220
AAAATGCACA	CCCTTTCT	GTGGCAAATC	GTCTCTGTAC	ATGTGTGTC	ATATTAGAAA	2280
GGGAAGATGT	TAAGATATGT	GGCCTGTGGG	TTACACAGGG	TGCCTGCAGC	GGTAATATAT	2340
TTTAGAAATA	ATATATCAA	TAACTCAACT	AACCTCAATT	TTAACATCA	TATTAATT	2400
TTTTCTTT	TAAAGAGAAA	GCAGGCTTT	CTAGACTTTA	AAGATAAAAG	TCTTGGGAG	2460
GTCTCACCGGT	GTAGAGAGGA	GCTTGAGGG	CCCCGCA	AAATTCAACCC	AGAGGGAAAT	2520
CTCGTCGGAA	GGACACTCAC	GGCAGTTCTG	GATCACCTGT	GTATGTCAAC	AGAAGGGATA	2580
CCGTCTCCCT	GAAGAGGGAA	CTCTGCACT	CCTCATGCCT	GTCTAGCTA	TACACCCATT	2640
TCTCTTGTCT	TCACAGGTT	TAAACTGGT	TTTGCAATAC	TGCTATATAA	TTCTCTGTCT	2700
CTCTCTGTCT	ATCTCTCCCC	TCCCTCCCT	CCCCCTCTTC	TCCATCTCCA	TTCTTTGAA	2760
TTTCCATCTC	CCTCCATCTC	AATCCCGTAT	CTACGCACCC	CCCCCCCCCCC	AGGCAAAGCA	2820
GTGCTCTGAG	TATCACATCA	CACAAAGGA	ACAAAAGCGA	AACACACAA	CCAGCCTCAA	2880
CTTACACTTG	GTACTCAAA	AGAACAAAGAG	TCAATGGTAC	TTGTCCTAG	GTGTTGGAAG	2940
AGGAAAACAG	GAACCCACCA	AACCAACAA	TCAACCAAAC	AAAGAAAAAA	TTCCACAATG	3000
AAAGAATGTA	TTTGTCTT	TTGCATTGT	GTGTATAAGC	CATCAATATT	CAGCAAAATG	3060
ATTCCCTTCT	TTAAAAAAA	AAATGTGGAG	GAAAGTAGAA	ATTTACCAAG	GTTGTTGGCC	3120
CAGGGCGTTA	AATTACAGA	TTTTTTAAC	GAGAAAAAAC	CACAGAAGAA	GCTACCTCAG	3180
GTGTTTTAC	CTCAGCACCT	TGCTCTGTG	TTCCCTTAG	AGATTTGTG	AAGCTGATAG	3240
TTGGAGCATT	TTTTTATT	TTAATAAAA	ATGAGTTGGA	AAAAAAATAA	GATATCAACT	3300
GCCAGCGCTGG	AGAAGGTGAC	AGTCAAAGTG	TGCAACAGCT	GTTCTGAATT	GTCTCCGCT	3360
AGCCAAGAAC	CNATATGGCC	TTCTTTGGA	CAAACCTGAA	AAATGTTAT	TT	3412

<210> 7
 <211> 1946
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 7

GCTGTAGCGG	AGGGGCTGGG	GGGCTGCTCT	GTCCCCTTCC	TTGCGCGCTG	CGGCCCTCAGC	60
CCACCCAGAG	GCCGGGGTGG	GAGGGCGAGT	GCTCAGCTTC	CCGGGTTAGG	AGCCGGAAA	120
TTCAAAATCCG	AAATATTCCA	CCCCAGCTCC	GATGGGAAGT	ACTGGACAGC	CTGCTGGCTC	180
AGTATGGTAC	AGTAGAGAAC	TGTGAGCAAG	TGAACACCGA	GAGTGAGACG	GCAGTGGTGA	240
ATGTACACCTA	TTCCAACCGG	GAGCAGACCA	GGCAAGCCAT	CATGAAGCTG	AATGCCACC	300
AGTTGGAGAA	CCATGCCCTG	AAGGTCCTCT	ACATCCCCGA	TGAGCAGATA	GCACAGGGAC	360
CTGAGAATGG	GCGCCGAGGG	GGCTTGGCT	CTCGGGGTCA	GGCCCGCCAG	GGCTCACCTG	420
TGGCAGCGGG	GGCCCCAGCC	AAGCAGCAGC	AAAGTGGACAT	CCCCCTTCGG	CTCCTGGTGC	480
CCACCCAGTA	TGTGGGTGCC	ATTATTGGCA	AGGAGGGGGC	CACCATCCGC	AACATCACAA	540
AACAGACCCA	GTCCAAGATA	GACGTGCATA	GGAGGAGAGA	CGCAGGTGCA	GCTGAAAAAG	600
CCATCAGTGT	GCACCTCCACC	CCTGAGGGCT	GCTCCCTCCGC	TTGTAAGATG	ATCTTGGAGA	660
TTATGCATAA	AGAGGCTAAAG	GACACAAAAA	CCGCTGACGA	GGTTCCTCTG	AAGATCCTGG	720
CCCATAATAA	CTTGTAGGG	CGTCTCATTC	GCAAGGAAGG	ACGGAACTCG	AAGAAGCTAG	780
AGCAAGATAC	CGAGACAAAAA	ATCACCATCT	CCTCGTTGCA	AGACCTTACC	CTTTACAACC	840
CTGAGAGGAC	CATCACTGTG	AAAGGGGCCA	TCGAGAATTG	TTGCAGGGCC	GAGCAGGAAA	900
TAATGAAGAA	AGTTCGGGAG	GCCTATGAGA	ATGATGTGGC	TGCCATGAGC	TCTCACCTGA	960
TCCCTGGCCT	GAACCTGGCT	GCTGTAGGTC	TTTCTCCAGC	TTCATCCAGC	GCAGTCCCGC	1020
CGCCTCCAG	CAGCGTTACT	GGGGCTGCTC	CCTATAGCTC	CTTATGCAG	GCTCCCGAGC	1080
AGGAGATGGT	GCAGGTGTTT	ATCCCCGCC	AGGCAGTGGG	CGCCATCATC	GGCAAGAAGG	1140
GGCAGCACAT	CAAACAGCTC	TCCCAGTTG	CCAGCGCCTC	CATCAAGATT	GCACCAACCG	1200
AAACACCTGA	CTCCAAAGTT	CGTATGGTA	TCATCACTGG	ACCGCCAGAG	GCCCAATTCA	1260
AGGCTCAGGG	AAAGATCTAT	GGCAAACCTCA	AGGAGGGAGA	CTTCTTTGGT	CCCAAGGAGG	1320
AAGTGAAGCT	GGAGACCCCAC	ATACGTGTGC	CAGCATCAGC	AGCTGGCCGG	GTCATTGGCA	1380
AAGGTGGAAA	AACGGTGAAC	GAGTTGCAGA	ATTGACGGC	AGCTGAGGTG	GTAGTACCAA	1440
GAGACCAGAC	CCCTGATGAG	AACGACCAGG	TCATCGTGAA	AATCATCGGA	CATTTCTATG	1500
CCAGTCAGAT	GGCTCAACCG	AAGATCCGAG	ACATCCCTGGC	CCAGGTTAACG	CAGCAGCATC	1560
AGAAGGGACA	GAGTAACCCAG	GCCCAGGCAC	GGAGGAAGTG	ACCAGCCCT	CCCTGTCCCT	1620
TNGAGTCCAG	GACAACAAACG	GGCAGAAATC	GAGAGTGTGC	TCTCCCCGGC	AGGCCTGAGA	1680
ATGAGTGGGA	ATCCGGGACA	CNTGGGGCG	GCTGTAGATC	AGGTTGCCC	ACTTGATTGA	1740
GAAAGATGTT	CCAGTGAGGA	ACCTGTATCT	NTCAGCCCCA	AACACCCACC	CAATTGGCCC	1800
AACACTGTNT	GCCCTCTGGG	GTGTCAGAAA	TTNTAGCGCA	AGGCACTTT	AAACGTGGAT	1860
TGTTAAAGA	AGCTCTCCAG	GCCCCACCAA	GAGGGTGGAT	CACACCTCAG	TGGGAAGAAA	1920
AATAAAATTT	CCTTCAGGTT	TTAAAAA				1946

<210> 8
 <211> 3283
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 8

GGCAGCGGAG	GAGGGGAGGA	GCGCCGGTAA	CCGGGGCCGGG	GGAGCCGCCG	GCTCTCGGGG	60
AAGAGACGGA	TGATGAACAA	GCTTTACATC	GGGAACCTGA	GCCCCGCCGT	CACCGCCGAC	120
GACCTCCGGC	AGCTCTTGG	GGACAGGAAG	CTGCCCCCTGG	CGGGACAGGT	CCTGCTGAAG	180
TCCGGCTACG	CCTTCGTGGA	CTACCCCGAC	CAGAACTGGG	CCATCCCGC	CATCGAGACC	240
CTCTCGGGTA	AACTGGAAATT	GCATGGAAA	ATCATGGAAAG	TTGATTACTC	ACTCTCTAAA	300
AAGCTAAGGA	GCAGGAAAT	TCAGATTGCA	AACATCCCTC	CTCACCTGCA	GTGGGAGGTG	360
TTGGATGGAC	TTTTGGCTCA	ATATGGACA	GTGGAGAATG	TGGAACAAGT	CAACACAGAC	420
ACAGAAACCG	CCGTTGTCAA	CGTCACATAT	GCAACAAGAG	AAGAAGCAA	AATAGCCATG	480
GAGAAGCTAA	GCGGGCATCA	GTTTGAGAAC	TACTCCTTCA	AGATTCCTA	CATCCGGAT	540
GAAGAGGTGA	GCTCCCCCTTC	GGCCCTCTAG	CGAGCCCAGC	GTGGGGACCA	CTCTTCCCGG	600
GAGCAAGGCC	ACGCCCCCTGG	GGGCACTTCT	CAGGCCAGAC	AGATTGATT	CCCGCTGCGG	660
ATCCTGGTCC	CCACCCAGTT	TGTTGGTGC	ATCATCGGAA	AGGAGGGCTT	GACCATAAAG	720
AACATCACTA	AGCAGACCCA	GTCCCCGGTA	GATATCCATA	AAAAAGAGAA	CTCTGGAGCT	780
GCAGAGAAGC	CTGTCACCAT	CCATGCCACC	CCAGAGGGGA	CTTCTGAAGC	ATGCCGCATG	840
ATTCTTGAAA	TCATGCAGAA	AGAGGCAGAT	GAGACCAAAC	TAGCCGAAGA	GATTCCCTCTG	900
AAAATCTTGG	CACACAATGG	CTTGGTTGGA	AGACTGATTG	AAAAAGAAGG	CAGAAATTG	960

AAGAAAAATTG AACATGAAAC AGGGACCAAG ATAACAATCT CATCTTGCA GGATTTGAGC 1020
 ATATACAACC CGGAAAGAAC CATCACTGTG AAGGGCACAG TTGAGGCCTG TGCCAGTGCT 1080
 GAGATAGAGA TTATGAAGAA GCTGCGTGAG GCCTTTGAAA ATGATATGCT GGCTGTTAAC 1140
 ACCCACTCCG GATACTTCTC CAGCCTGTAC CCCCATCACC AGTTTGGCCC GTTCCCGCAT 1200
 CATCACTCTT ATCCAGAGCA GGAGATTGTG AATCTCTTCA TCCCAACCCA GGCTGTGGGC 1260
 GCCATCATCG GGAAGAAGGG GGCACACATC AAACAGCTGG CGAGATTCGC CGGAGCCTCT 1320
 ATCAAGATTG CCCCTGCGGA AGGCCCAGAC GTCAAGCGAAA GGATGGTCAT CATCACCGGG 1380
 CCACCGGAAG CCCAGTTCAA GGCCCAGGGG CGGATCTTTG GGAAACTGAA AGAGGAAAAC 1440
 TTCTTAACC CCAAAGAAGA AGTGAAGCTG GAAGCGCATA TCAGAGTGCCTCTCCACA 1500
 GCTGCCGGG TGATTGGCAA AGGTGGCAAG ACCGTGAACG AACTGCAGAA CTTAACCAAGT 1560
 GCAGAAGTCA TCGTGCCTCG TGACCAAACG CCAGATGAAA ATGAGGAAGT GATCGTCAGA 1620
 ATTATCGGGC ACTTCTTTGC TAGCCAGACT GCACAGCGCA AGATCAGGGA AATTGTACAA 1680
 CAGGTGAAGC AGCAGGAGCA GAAATACCC CAGGGAGTCG CCTCACAGCG CAGCAAGTGA 1740
 GGCTCCCAACA GGCACCAGCA AAACAAACGG AATGTAGC CTTCCAACA CCTGACAGAA 1800
 TGAGACAAA CGCAGCCAGC CAGATCGGG ACAAACCAAA GACCATCTGA GGAATGAGAA 1860
 GTCTCGGGAG GCGGCCAGGG ACTCTGCCGA GGCCCCTGAGA ACCCCCAGGGG CCGAGGAGGG 1920
 GCGGGGAAGG TCAGCCAGGT TTGCCAGAAC CACCGAGGCC CGCCTCCCGC CCCCCAGGGC 1980
 TTCTGCAGGC TTCAAGCCATC CACTTCACCA TCCACTCGGA TCTCTCTGA ACTCCACGA 2040
 CGCTATCCCT TTAGTTGAA CTAACATAGG TGAACGTGTT CAAAGCCAAG CAAAATGCAC 2100
 ACCCTTTTC TGTGGCAAAT CGTCTCTGTA CATGTGTGTA CATATTAGAA AGGGAAAGATG 2160
 TTAAGATATG TGGCCTGTGG GTTACACAGG GTGCCCTGCAG CGGTAATATA TTTTAGAAAT 2220
 AATATATCAA ATAACCTAAC TAACTCCAAT TTTTAATCAA TTATTAATTT TTTTTCTTT 2280
 TTAAGAGAA AGCAGGCTTT TCTAGACTTT AAAGAATAAA GTCTTGAGG GGTCTCACGG 2340
 TGTAGAGAGG AGCTTTGAGG CCACCCGCAC AAAATTCAAC CAGAGGGAAA TCTCGTCGGA 2400
 AGGACACTCA CGGCAGTTCT GGATCACCTG TGTATGTCAA CAGAAGGGAT ACCGCTCTCCT 2460
 TGAAGAGGAA ACTCTGTCAAC TCCTCATGCC TGTCTAGCTC ATACACCCAT TTCTCTTGC 2520
 TTCACAGGTT TTAAACTGGT TTTTGCCATA CTGCTATATA ATTCTCTGTC TCTCTCTGTT 2580
 TATCTCTCCC CTCCCTCCCC TCCCCCTCTT CTCCATCTCC ATTCTTTGA ATTCTCTCAT 2640
 CCCTCCATCT CAATCCCCGT A TCTACCGCACC CCCCCCCCCCCC CAGGCAAAGC AGTGTCTGTA 2700
 GTATCACATC ACACAAAAGG AACAAAAGCG AACACACACAA ACCAGCCCTCA ACTTACACTT 2760
 GGTTACTCAA AAGAACAAAGA GTCAATGGTA CTGTCCTAG CGTTTGAGA GAGGAAAACA 2820
 GGAACCCACC AAACCAACCA ATCAACCAA CAAAGAAAAA ATTCCACAAAT GAAAGAATGT 2880
 ATTTGTCTT TTGCACTTTT GGTGTATAAG CCATCAATAT TCAGCAAAT GATTCTTTTC 2940
 TTTAAAAAAA AAAATGTGGA GGAAAGTAGA AATTTACCAA GGTTGTTGGC CCAGGGCGTT 3000
 AAATTACACAG ATTTTTTAA CGAGAAAAAC ACACAGAAGA AGCTACCTCA GGTGTTTTA 3060
 CCTCAGCACC TTGCTCTTGT GTTCCCTTA GAGATTTGT AAAGCTGATA GTTGGAGCAT 3120
 TTTTTTATTT TTTAATAAA ATGAGTTGG AAAAAAAAATA AGATATCAAC TGCCAGCCTG 3180
 GAGAAGGTGA CAGTCCAAGT GTGCAACAGC TGTCTGAAT TGTCTCCGC TAGCCAAGAA 3240
 CCNATATGGC CTTCTTTGG ACAAACCTTG AAAATGTTA TTT 3283

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/05766

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G01N 33/574, 33/48; C12Q 1/68
 US CL : 435/7.23, 6; 436/64

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.23, 6; 436/64

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TURECI, O. et al. Expression of SSX Genes in Human Tumors International Journal of Cancer. 1998, Vol. 77, No. 1, pages 19-23, especially page 19.	41-52
A	GURE, A.O. et al. SSX: A Multigene Family with Several Members Transcribed in Normal Testis and Human Cancer. International Journal of Cancer. 1997, Vol. 72, No. 6, pages 965-971, especially page 965.	41-52
A	CHEN, et al. A Testicular Antigen Aberrantly Ex[pressed] in Human Cancers Detected by Autologous Antibody Screening. Proceedings of the National Academy of Science. March 1997, Vol. 94, No. 5, pages 1914-1918, especially page 1914.	41-52

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance		
E earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
O document referring to an oral disclosure, use, exhibition or other means	"A"	document member of the same patent family
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

21 JULY 1999

Date of mailing of the international search report

11 AUG 1999

Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

YVONNE EYLER

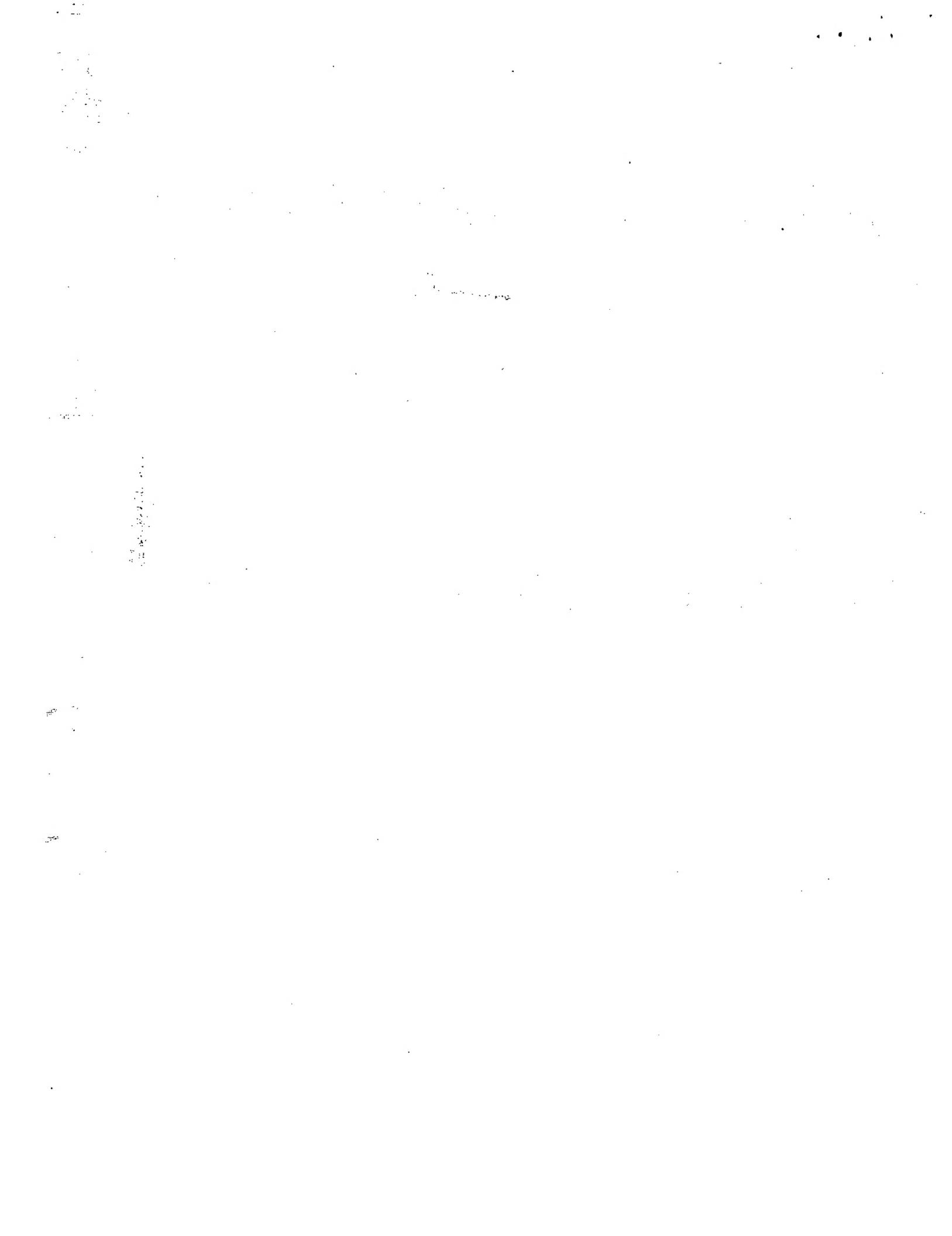
Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/05766

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,Y	CHEN, et al. Identification of Multiple Cancer/Testis Antigens by Allogeneic Antibody Screening of a Melanoma Cell Line Library." Proceedings of the National Academy of Science. June 1998, Vol. 95, No. 12, pages 6919-6923, especially page 6919.	41-52
P,Y	UGUR, S. et al. Expression of Multiple Cancer/Testis (CT) Antigens in Breast Cancer and Melanoma: Basis for Polyvalent CT Vaccine Strategies. International Journal of Cancer. 1998, Vol. 78, No. 3, pages 387-389, especially page 387.	41-52



<110> Chen, Yao-Tseng
Curc, Ali.
Tsang, Salem
Stockert, Elisabeth
Jager, Elke
Knuth, Alexander
Old, Lloyd J.

<120> Isolated Nucleic Acid Molecules Encoding Cancer Associated Antigens, The Antigens Per Se, And Uses Thereof

<130> LUD 5538.1 PCT

<140> PCT/US 99/05766

<141> 1998 - 04 - 17

<160> 3

<210> 1
<211> 4265
<212> DNA
<213> Homo sapiens
<220>
<400> 1

GTCTGAGGAA	CCTGAGGCAT	TTTCTGACCA	GGATCCTCTC	AGGTCAAGCGG	AGGGAGGAGA	60
CTTATAGACC	TATCCAGTCT	TCAACGCTCT	CCAGGAAAGCA	GGAGTTGAAG	ACCTGGGTGT	120
GAGGGACACA	TACATTTCTTA	AAGCACCACG	CCAGGAGGGG	CCCAGACAGT	GGCAGGAGTC	180
AAGGTTCCCA	GGAGACAAAC	CCCTTAGGAA	CCACAGGCGAC	CTCTGAGGCC	CTAGAGLACC	240
ACCTTAAGAG	AAGAAGAGCT	CTAACCGCCG	CTTTCGACAGA	CCCATCATGG	GGGACAAAGGA	300
TATGCCACT	CCTCGGATGC	CCAGCTCTCT	CCAGAGTTCC	TCTGAGAGTC	CTCAAGAGTG	360
TCCGTAGGGG	GAGGACTCCC	AGTCTCTCTC	CCAGATTCCT	CTGAGTTCTC	CTGAGAGCGA	420
CGACACCCCTG	TATCCCTCTCA	AGATCTCTCA	GAATCGTTCT	CTGGGGGAGG	ACTCTCGGA	480
TCCCTCTCCAG	AGACCTCTCTC	AGGCGAGGAA	CTCCCCAGCT	CTCTCTCCAGA	TTCCCCAGAG	540
TTCTCTCTGAG	GGUGAOGAGC	CCCAGTCTCC	TCTCCAGAAAT	TCTGAGAGTT	CTCTCTGAGGG	600
GANGGACTTC	CTCTCTCTTC	TAGAGMTTTC	TGAGGAGCCCT	TCTGAGGGTG	ACGATGTCGA	660
GTCTCCCTCTG	CAAGATUUTG	CGAGCTCCCT	CTTCCTCTCT	GTCTTATTGA	GTATTTTCCA	720
GAGPTCCCCCT	GAGNGTATTC	AAAGTCTCTT	TGAGGGTTTT	CCCCAGTCTG	TTCTCCAGAT	780
TOCTCTGAGC	CCCGCTCTCT	CCTCCACCTT	AGTGAAGTATT	TTCCAGAGTT	CCCCCTCAGAS	840
TACTCPAAGT	CCCTTTGAGG	GTTTTCCCCA	GTCTCCACTC	CAGATTCCTG	TGAGCCCCCTC	900
CTTCTCTCTCC	ACTTTTATTGA	GTATTCTCCA	GAATTCCTCT	GAGAGAAGTC	AGAGAACTTC	960
TGAGGGTTTT	CCACAGTCTC	CTCTCCAGAT	TCTCTGTGAGC	TCCTCTCTGT	CCCTCCTCTT	1020
ACTGAGTCTT	TTCCAGAGTT	CCCCCTGAGAG	AACTCTAGAT	ACTTTTGAGG	GTTTTCCCCA	1080
GTCTCCACTC	CAAGATCTCTG	TGAGCCGCTC	CTTCTCTCTC	ACITTTATTGA	GTATTTTCCA	1140
GAGTTCCCCCT	GAGAGAAGTC	AGAGTACTTT	TGAGGGTTTT	GCCCGAGTCTC	CTCTCCAGAT	1200
TCCGTGAGC	CCCTCTCTCT	CCTCACAGTT	ACTGAGTATT	TTCCAGAGTT	CCCTGAGAG	1260
AACTCTAGAT	ACTTTTCAAGG	TTTTCCTCCCT	CTCTCTCTCT	CAGATTCCTG	TAAGCTTCTC	1320
CTTCTCTCTCC	ACTTTTATTCA	GTCTTTTCCA	GAGTCTCTCT	GAGAGAAGTC	AGAGTACTTT	1380
TGAGGGTTTT	CCUCAGTCTC	CTCTCCAGAT	TCTCTGAGAG	CCCTCTCTCT	CCTOCACATT	1440
ACTGAGTCTT	TTCCAGAGTT	CCCCCTGAGAG	AACTCACAGT	ACTTTTGAGG	GTTTTCCCCA	1500
GTCTCCCTCTC	CAAGATCTCTG	CTTCTCTCTC	ACTTTATTGA	GTATTTTACA	1560	
GAGTTCTCTCT	GGAGAGTGCCTC	AAAGTCTTTT	TGAGGGTTTT	CCCCAGTCTC	CTCTCTCAGAT	1620
TCCGTGAGC	TCTCTCTCTC	CCTACACTTT	ATTGAGTCTT	TCTCAGAGTT	CCCCCTGAGAG	1680
AACTCACAGT	ACTTTTCAAGG	GTTTTCCCCA	GTCTCTCTCTC	CAGATTCCTG	TGAGCTCTC	1740
CTTCTCTCTCC	TCCACCTTAT	TGAGCTCTTT	CCAGAGTCTC	CTCTGAGTGA	CTCAAAGTAC	1800
TTTTRAGGGT	TTTTCCTCTG	CTCTCTCTCA	GAATCCCTCAG	ACTCTCTCTG	AAGGGAGAGA	1860
TACCCATTCT	CTCTCTCTCA	TTGTTCCCTAAG	TCTCTCTCTAG	TGGGAGGGCT	CCCTGCTCTC	1920
TCACTACTTT	CTCTCTCTGCC	CTCTCTCTGGG	GGAGGAGCTC	CTATCTCTC	ACTACTTTCC	1980
TCAGAGCCCT	CTCTCTCTGGG	AGGACTCTCT	CTCTCTCTCA	TACTTTCTC	AGAGCCCCCA	2040
GGGGGAGGAC	TCCCTGTCTC	CTCACTACTT	TCCTCAGAGC	CTCTCTCTGG	GGGAGGAGTC	2100
CATGTCCTCT	CTCTACTTTC	CTCAGAGTCC	TCTTCAGGGG	GGGAAATTCC	AGTCTTCTCT	2160
CCAGAGCCCT	GTGAGGATCT	GCTCTCTCTC	CACTCCATCC	AGTCTTCCCT	AGAGTTTCCC	2220
TGAGAGTTCT	CTGAGCTCTC	CTGAGGGGCC	TGTCAGCTCT	CTCTCTCAGAG	CTCTCTCAGAG	2280

GGCTCTGAG	GGGATGCAGT	CCAAATCTCC	TCTCCAGAGT	CTCTAGAGTG	CTCCAGGGG	2348
GGAGGATTCU	CTGTCCTCTC	TCCAAATTCCTC	TCAGAGTCCT	CTTCAAGAGAG	AGGAACTCCCF	2400
CTCTCTCTC	CAATTCTCTC	AGATCTCTCC	TGACTGGGAG	GAATCCCTCT	CTCCCTCTCA	2450
CTTTCCTCTAG	TTCCTCTCTC	AGGGGGAGGA	CITCCAGTCT	TCTCTCCAGA	CTOCTGTGAG	2520
TATCTGCTCC	TCCTCCACTT	CTTGAGICCT	TCCCCAGAGT	TTTCTGAGA	GTOCTCAGAG	2580
TCCCTCTGAG	GGGCCTCTCTC	AGTCCTCTCT	CCAGAGACCT	GTCAGCTCT	TCCFCTOCPA	2640
CACTTTAGCA	AGTCTCTCTCC	AAAGTTUCCA	TCAGAGTCCT	CAGAGTCCTC	CTGAGGGGCC	2700
TGCCCAAGTCT	CCTCTCCAGA	GTCTCTCTTG	CCTCTTCCCC	TCCTCCACTT	CTTCGAGTCT	2760
TTCCCCAGAGT	TCTCTCTGTA	CTCCCTTCCC	CTCTCTCCACT	TCATCGAGTC	TTTCCAGAG	2820
TTCCCCCTGAG	AGTCCTCTCC	AGAGTCCTGT	GATCTCTTTC	TCCTCTCTCA	CTTCATTGAG	2880
CCATTCTACT	GAAGAGTCGA	GCAGCCCCAGT	AGATGATAT	ACAAGTTCT	CAGACACCTT	2940
GTAGAGAGT	GATTCCTTGA	CAGACAGCGA	TCCTCTGATA	GAGAGGAGC	CTTGTTCAC	3000
TTATACACTG	GATGAAAAGG	TGGACCGAGT	GGCGCGTTT	CTTCCTCTCA	AAATATCAAGT	3060
GAAGCAGGCT	ATCACAAAGG	CAAGAGATGCT	GACGAAATGTC	ATCAAGLAGGT	ACRCGGGCTA	3120
CTTTOCTGTC	ATCTTCAGGA	AGCCCCGTTGA	GTTCTATAGTG	ATACITTTTG	GCATTTCCCT	3180
GAGAGGAGTG	GACCTCTATG	ACTCTCTATGT	CTTGTTAAC	ACATTAGACG	TCBCCTCTGA	3240
GGGGTCTCTG	AGTGAATGAGC	AGGECATATGTC	CCAGAACCGC	CTCTCTGATTC	TTATTCAG	3300
TATCATCTTC	ATAAAAGGGCA	CCTATGCTTC	TGAGGGGGTC	ATCTGGGATG	TCCTCAGTGG	3360
AAATGGGGTG	CTGCTGGGGG	GGEAGACACT	TGCTCTTGGG	GGCCTTGGG	AGCCTCTCAC	3420
TAAGATTTGG	CTGCAAGGAAAC	TTTACCTTGA	GTACCCGGAG	GTGCCCCAAT	CTTCCTCTOC	3480
TGCTTACGAA	TTCTCTCTGGG	GTCTCAAGAC	TCATTCAGA	GTCTTAAAGA	GGAAAGTAGT	3540
AGAGTTTTTG	GGCATGCTTA	AGATATACCGT	CCCTTATTAAC	TTTCTCTCTT	CTTACUAGGA	3600
TGCTTTGAAA	GATGTGGAAG	AGAGAGCCCCA	GGCCATATT	GACACCACAG	ATGATTGCG	3660
TOCCACAGAA	AATGCAAGCT	CAAGTCTCAT	GTCCCCCAGC	TTCTCTCTTG	ATGAACTCT	3720
AGGCCAGATT	CTTCCTCTC	AGTTTGAGG	GGGCAGTCGA	CTTCTCTACGT	GTGAGGGGCT	3780
CTGGTTGAGG	CTGGAGAGAA	CACAGTGCTA	TTTGCATTTC	TGTCTCATAT	GGGTAGTTAT	3840
GGGGTTTAC	TGTTTACTT	TTCGGTTATT	TTCAAAAGCT	TTTCTCTATTA	ATAACAGGTT	3900
TAATATGCTT	CAGAATCTTA	GTTTATGCA	ATGAGTCGCA	CATGTATTGC	TGTTTTCTG	3960
GTTTAAAGT	AACAGTTGAA	TATTTGTAA	AAACAAACAC	ACACCCAAAC	ACACCACATT	4020
GGGAACACCT	TCTGCTCTCAT	TTTGTGATCT	GTCAAGGTT	ATGTTGGTGT	TACTGTAGGA	4080
ATTTCTCTGA	AACTGTGAAAG	GAACCTCTGCA	CTTAAATAGT	GGAAATAAAGT	AAAGGATTCT	4140
TAATGTGTTG	ATTTCTCTAG	GTCTTTAGT	CTGTTGTTCT	TGAAAACAA	AGATACATAC	4200
CTGGTTTGCT	TGGCTTACGT	AAAGAAAGTC	AAAGAAAGTAA	ACTGTAATAA	ATAAAACCT	4260
CAGTG						4265

<210> 2
<211> 1142
<212> PRT
<213> Homo sapiens
<220>
<400> 2

Met	Gly	Asp	Lys	Asp	Met	Pro	Thr	Ala	Gly	Met	Pro	Ser	Leu	Leu	Gln		
								5						10			15
Ser	Ser	Ser	Glu	Ser	Pro	Gln	Ser	Cys	Pro	Glu	Gly	Glu	Asp	Ser	Gln		
								20						25			30
Ser	Pro	Leu	Gln	Ile	Pro	Gln	Ser	Ser	Pro	Glu	Ser	Asp	Asp	Thr	Leu		
								35						40			45
Tyr	Pro	Leu	Gln	Ser	Pro	Gln	Ser	Arg	Ser	Glu	Gly	Glu	Asp	Ser	Ser		
								50						55			60
Asp	Pro	Leu	Gln	Arg	Pro	Pro	Glu	Gly	Lys	Asp	Ser	Gln	Ser	Pro	Leu		
								65						70			75
Gln	Ile	Pro	Gln	Ser	Ser	Pro	Glu	Gly	Asp	Asp	Thr	Gln	Ser	Pro	Leu		
								85						90			95

Gln	Asn	Ser	Gln	Ser	Ser	Pro	Glu	Gly	Lys	Asp	Ser	Leu	Ser	Pro	Leu
			100				105					110			
Glu	Ile	Ser	Gln	Ser	Pro	Pro	Glu	Gly	Glu	Asp	Val	Gln	Ser	Pro	Leu
			115				120					125			
Gln	Asn	Pro	Ala	Ser	Ser	Phe	Ph	Ser	Ser	Ala	Leu	Leu	Ser	Ile	Phe
			130			135					140				
Gln	Ser	Ser	Pro	Glu	Ser	Ile	Gln	Ser	Pro	Phe	Glu	Gly	Phe	Pro	Gln
			145			150				155					160
Ser	Val	Leu	Gln	Ile	Pro	Val	Ser	Ala	Ala	Ser	Ser	Ser	Thr	Leu	Val
					165					170					175

Ser Ile Phe Gln Ser Ser Pro Glu Ser Thr Gln Ser Pro Phe Gln Gly
 180 185 190
 Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser Arg Ser Phe Ser Ser
 195 200 205
 Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro Gln Arg Ser Gln Arg Thr
 210 215 220
 Ser Glu Gly Phe Ala Gln Ser Pro Leu Gln Ile Pro Val Ser Ser Ser
 225 230 235 240
 Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr
 245 250 255
 Gln Ser Thr Phe Gln Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val
 260 265 270
 Ser Arg Ser Phe Ser Ser Thr Leu Leu Ser Ile Phe Gln Ser Ser Pro
 275 280 285
 Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Ala Gln Ser Pro Leu Gln
 290 295 300
 Ile Pro Val Ser Pro Ser Phe Ser Ser Thr Leu Val Ser Ile Phe Gln
 305 310 315 320
 Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser
 325 330 335
 Pro Leu Gln Ile Pro Val Ser Ser Phe Ser Ser Thr Leu Leu Ser
 340 345 350
 Leu Phe Gln Ser Ser Pro Glu Arg Thr Gln Ser Thr Phe Gln Gly Phe
 355 360 365
 Pro Gln Ser Pro Leu Gln Ile Pro Gly Ser Pro Ser Phe Ser Ser Thr
 370 375 380
 Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu Arg Thr His Ser Thr Phe
 385 390 395 400
 Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Met Thr Ser Ser Phe
 405 410 415
 Ser Ser Thr Leu Leu Ser Ile Leu Gln Ser Ser Pro Glu Ser Ala Gln
 420 425 430
 Ser Ala Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile Pro Val Ser
 435 440 445
 Ser Ser Phe Ser Tyr Thr Leu Leu Ser Leu Phe Gln Ser Ser Pro Glu
 450 455 460
 Arg Thr His Ser Thr Phe Glu Gly Phe Pro Gln Ser Pro Leu Gln Ile
 465 470 475 480
 Pro Val Ser Ser Ser Ser Ser Thr Leu Leu Ser Leu Phe Gln
 485 490 495
 Ser Ser Pro Gln Cys Thr Gln Ser Thr Phe Glu Gly Phe Pro Gln Ser
 500 505 510
 Pro Leu Gln Ile Pro Gln Ser Pro Pro Gln Gly Glu Asn Thr His Ser
 515 520 525
 Pro Leu Gln Ile Val Pro Ser Leu Pro Gln Trp Glu Asp Ser Leu Ser
 530 535 540
 Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Leu Ser
 545 550 555 560
 Pro His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Gln Asp Ser Leu Ser
 565 570 575
 Pro His Tyr Phe Pro Gln Ser Pro Gln Gly Glu Asp Ser Leu Ser Pro
 580 585 590
 His Tyr Phe Pro Gln Ser Pro Pro Gln Gly Glu Asp Ser Met Ser Pro
 595 600 605
 Leu Tyr Phe Pro Gln Ser Pro Leu Gln Gly Glu Glu Phe Gln Ser Ser
 610 615 620
 Leu Gln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Pro Ser Ser Leu
 625 630 635 640
 Pro Gln Ser Phe Pro Gln Ser Ser Gln Ser Pro Pro Gln Gly Pro Val
 645 650 655
 Gln Ser Pro Leu His Ser Ser Pro Gln Ser Pro Pro Gln Gly Met His Ser
 660 665 670
 Gln Ser Pro Leu Gln Ser Pro Gln Ser Ala Pro Gln Gly Glu Asp Ser
 675 680 685

Leu Ser Pro Leu Gln Ile Pro Gln Ser Pro Leu Glu Gly Glu Asp Ser
 690 695 700
 Leu S r Ser Leu His Phe Pro Gln Ser Pro Pro Glu Trp Glu Asp Ser
 705 710 715 720
 Leu Ser Pro Leu His Phe Pro Gln Phe Pro Pro Gln Gly Glu Asp Phe
 725 730 735
 Gln Ser Ser Leu Cln Ser Pro Val Ser Ile Cys Ser Ser Ser Thr Ser
 740 745 750
 Leu Ser Leu Pro Gln Ser Phe Pro Glu Ser Pro Gln Ser Pro Pro Glu
 755 760 765
 Gly Pro Ala Gln Ser Pro Leu Gln Arg Pro Val Ser Ser Phe Phe Ser
 770 775 780
 Tyr Thr Leu Ala Ser Leu Leu Gln Ser Ser His Glu Ser Pro Gln Ser
 785 790 795 800
 Pro Pro Glu Gly Pro Ala Gln Ser Pro Leu Gln Ser Pro Val Ser Ser
 805 810 815
 Phe Pro Ser Ser Thr Ser Ser Leu Ser Gln Ser Ser Pro Val Ser
 820 825 830
 Ser Phe Pro Ser Ser Thr Ser Ser Leu Ser Lys Ser Ser Pro Glu
 835 840 845
 Ser Pro Leu Gln Ser Pro Val Ile Ser Phe Ser Ser Ser Thr Ser Leu
 850 855 860
 Ser Pro Phe Ser Glu Glu Ser Ser Ser Pro Val Asp Glu Tyr Thr Ser
 865 870 875 880
 Ser Ser Asp Thr Leu Leu Glu Ser Asp Ser Leu Thr Asp Ser Glu Ser
 885 890 895
 Leu Ile Glu Ser Glu Pro Leu Phe Thr Tyr Thr Leu Asp Glu Lys Val
 900 905 910
 Asp Glu Leu Ala Arg Phe Leu Leu Lys Tyr Glu Val Lys Gln Pro
 915 920 925
 Ile Thr Lys Ala Glu Met Leu Thr Asn Val Ile Ser Arg Tyr Thr Gly
 930 935 940
 Tyr Phe Pro Val Ile Phe Arg Lys Ala Arg Glu Phe Ile Glu Ile Leu
 945 950 955 960
 Phe Cys Ile Ser Leu Arg Glu Val Asp Pro Asp Asp Ser Tyr Val Phe
 965 970 975
 Val Asn Thr Leu Asp Leu Thr Ser Glu Gly Cys Leu Ser Asp Gln Gln
 980 985 990
 Gly Met Ser Cln Asn Arg Leu Leu Leu Ile Leu Ser Ile Ile Phe
 995 1000 1005
 Ile Lys Gly Thr Tyr Ala Ser Glu Glu Val Ile Trp Asp Val Leu Ser
 1010 1015 1020
 Gly Ile Gly Val Arg Ala Gly Arg Glu His Phe Ala Phe Gly Glu Pro
 1025 1030 1035 1040
 Arg Glu Leu Leu Thr Lys Val Trp Val Cln Glu His Tyr Leu Cln Tyr
 1045 1050 1055
 Arg Glu Val Pro Asn Ser Ser Pro Pro Arg Tyr Glu Phe Leu Trp Gly
 1060 1065 1070
 Pro Arg Ala His Ser Glu Val Ile Lys Arg Lys Val Val Glu Phe Leu
 1075 1080 1085
 Ala Met Leu Lys Asn Thr Val Pro Ile Thr Phe Pro Ser Ser Tyr Lys
 1090 1095 1100
 Asp Ala Leu Lys Asp Val Cln Glu Arg Ala Gln Ala Ile Ile Asp Thr
 1105 1110 1115 1120
 Thr Asp Asp Ser Thr Ala Thr Cln Ser Ala Ser Ser Ser Val Met Ser
 1125 1130 1135
 Pro Ser Phe Ser Ser Glu
 1140

<210> 3
 <211> 7
 <212> PRT
 <213> Homo sapiens
 <220>
 <400> 3

Pro Glu Ser Pro Leu Glu Ile
 1 5

<210> 4
 <211> 4159
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 4

GG1GGATGCC TTTGGGTTGT AGCTAGGCTT TTTCTTTCTT TTCTCTTTTA AAACCATCT
 AGACAGGGAA AAAACAGGCG TUGGATCTGA TTTTCACTC CTGCTCTTG TGCTTCCTTC 60
 TTACCTGTGTT TCTGTATTTT AAAGGGCGAGA AGACGAGGCG ARCAAAACCA GCTGGATCCA 120
 TCCATCAGCG TCGGTGGTTT TAATTTTCC TTTTTCTCG TTATTTTTT TAAACAAACC 180
 ACTCTTCACA ATGAACRAAC TGTATATCGC ARACCTCAGC GAGAACGGCG CCGCTCTCGGA 240
 CCTAGGAGT ATCTTCAGG AGCGGAGAT CCCGGTGTGCG CGACCCCTTCC TGGTGAAGAC 300
 TGGCTACCGG TCGCTGGACT CGCUGGAGGA GAGCTGGGCC CTCTAGGCGA TCGAGGCGCT 360
 TTCAAGCTAA ATAGCTCTGC AGCGGAGAAC CATAGAAGT GAGCTCTCGG TCGCTAAAG 420
 GCAAGGATT CGGAAACTC AGATACGAA TATCCCGCT CATTTCAGT GGGAGGTGCT 480
 GGATAGTTA CTAGTCACT ATGGAGTGGT GGAGAGCTGT GAGCAAGTCA ACAGTGCCTC 540
 GGAACCTGCA GTTGTAAATG TAACCTTACG CAGTAAAGAC CGAGCTAGAC AAGCACTAGA 600
 CAACCTGAT CGATTTCTAGT TAGAGATTG CTCCTTGAAT CTAGCCATTA TCGCTGATGA 660
 ATATGCCGCG CTCGAAACCC CCTTCAGGCA CGCCCGAGGT CGCCGGGGGGC TTGGCGAGAG 720
 GGGCTCTCA AGGCAGGGGT CTCCAGGATC CGTATCCAAG CAGAACCCAT GTGATTTGCC 780
 TCTGGCCCTG CTGGTTUCCG CCGAATTGTTG TGGAGCCATC ATAGGAAAAG AAGGTGCCAC 840
 CATCGGAAC ATCACTAAAC AGACCCAGTC TAAATCGAT GTCCACCGTA AAGAAARTGC 900
 GGGGGCTCCT GAGNAGTGA TTACTATCCT CTCTACTCCT GAGCCACCT CTGCGCTTC 960
 TRAGTCTATT CTGGAGATTA TGCATAAGGA AGCTCAGAT ATAAATTCGA CAGAAGAGAT 1020
 CGGCTGAG ATTTTAGCTC ATAAATACCT TGTGGAUGT CTTATTGCTA AAGAAGGAAAG 1080
 AAATCTTAA AAGATTCAGC AAGACACAGA CAGTAAATC AGCATATCTC CATTGCAAGA 1140
 ATTGACGCTG TTTATCTCG AGCCGACTAT TACAGTTAAA GGCATAGTC AGACATGTC 1200
 CAAACCTGAG GAGGAGATCA TAAAGAATAT CAGGGAGTCT TATGAAAATG AGATTGCTTC 1260
 TATGAACTCTT CGAACACATT TAATTCCTGG ATAAATCTC PACGCGTCTGG CTCTGTTECC 1320
 ACCCACTTCA GGGATGCCAC CTCCACCTC AGGGCCCCCT TCAGOCATGA CTCCCTCCCTA 1380
 CCCCCAGCTG GAGCAATCAG AGACGGAGAC TCTTCATCAG TTTATCCAG CTCTATCAGT 1440
 CGGTGCCATC ATGGCCAGGC AGGGCCAGCA CATCAAGGAG CTTTCTCGCT TTGCTGGAGC 1500
 TTCAATTAGG ATTGCTCCAG CGGAAGCACC AGATGCTAAA GTGAGGATGG TGATTATCAG 1560
 TGGACCACCA EAGGCTCAGT TCAAGGCTCA GGGAAAGAATT TATGGAAAAA TTAAAGAAGA 1620
 AAACCTTGTGTT AGTCCTAAAC AAGAGGTGAA ACTTGAAGCT CATATCAGAG TGCCATCCTT 1680
 TGCTGCTGGC AGAGTTATG GAAAGGAGG CAAACCGGTG ATGAACTTC AGAATTGTC 1740
 AAGTGCAGAA GTTGGTGTCC CTCTGTGACCA GACACCTGAT GAGAATGACU AAGTCGGTGT 1800
 CAAATAAATCT CGTCAACTCT ATGCTGCGA CCTTCGCCCAG AGAAAATTC AGGAAATTCT 1860
 GACTCAGGTA AGACGACACC AACACAGAA CGCTCTGCGA AGTGGACCAU CTCAGTCAG 1920
 ACGGAAGTAA AGGCTCAAGA AAGCAGTCAC CACAGAGGCA GATGCCAAC CAAAGCAGA 1980
 TTGCTTAAAC AACAGATGGG CGCTGACCC CTATCCAGAA TCACTGCGAC AGTTTTTAC 2040
 CTAGCCCTGTT GTTCTGAGG ACCAGGCGAC TTTCAGCTC CTGCTCTGT GAGAATGTTAT 2100
 ACTTTATGCT CTCTGAAATC TATGACACCC AGCTTTAAA CAACAAACCA ACAAACCAA 2160
 AAGAAGGCTGG GGGAGGGAGG GAAAGAGAG AGCTCTGCGAC TTOCCTTGT TGTATGCTCA 2220
 CAGTATAACA GATATTCTAA TTCTCTTAA TATTCCTTAA TATGCGAGA AATGGCTTA 2280
 ATGATGCTTT CTCATCTC ATCAATAGA TTGCTCTTAA ATCCAAATGT TAAATTTGGA 2340
 TCACTTATCT TATGACAGGA ACTTAAATGT TAAAGCCATTA GCAAGAAGAA ACTGTTCTCA 2400
 GTTTTATTTT TACCTAACAC TAACTGAGT AACCTAAGGG AAGTGCAGA TGCTGTTGGC 2460
 AGGGGTATTA AACGTGCATT TTACTCAAC TACCTCAGGT ATTCACTTAAT ACAAAGAAA 2520
 GCAAAATTGT TCCCTTTTTT TGAAAATTAA ATATACTTTA TAAATGATAGA AGTCCAAACCG 2580
 TTTTTAAAA AATAATTAA AATTTAACCA GCAATCAGCT AACAGGCAA AATAGATTTT 2640
 TTTTTAAAA

TACTCTGGC TGGTGCACGT AAAGCTGGAA AATTAATTTC AGGGTTTTT GAGGCTTTG 2760
 ACACAGTTAT TACCTAAATC AAATGTCRA AATTAACGGAG CAGTGCCTAG TATCTGGAGA 2820
 GCGAGCACTAC CATTIATTCT TTCATTATA GTTGGGAAAG TTTTGACGG TACTAACAAA 2880
 GTGGTCGCGAG GAGATTTGG AACGGCTGGT TTAAATGGCT TCAGGAGACT TCAGTTTTT 2940
 GTTTCGCTAC ATGATTGAAAT GCATAATRAA TGCTTCTGTC TTCTGACTAT CTTACCTAA 3000
 AGARAGTGCA TCACTCAAGA GTGCGAGAC TTCTCACTGA CTGGCAAAA CGAAGTTTA 3060
 CCTTGTCTTA TAGGATGCTT AGTTGGCCAC TACCTTCAG ACCTATGGG CAGTCATAGA 3120
 TGGTGTGACA GTGTTAACCG CCAACAAAG GCTACATTC CAGGGGCCA CCAGCTGCT 3180
 GAGGCTCACT AAGCTATTTT CAAGATTTTT AAGGACTGAT AATTAAPAA AAAAAA 3240
 AATTTAGACT CCACCTTAAAG TAGTAAAGTA TACUACGATT TCTGTATACT CTGCACTAG 3300
 TTCTTGAA AAAAAAGTCRA AAGATAGAGA ATACAAAGAA AGTTTNTGGG ATATATTTG 3360
 ATGACTGTG AAAACATATG ACCTTTGATA ACGAACTCAT TTGCTCCTC CTTGACACCA 3420
 AAGCCCGATA CGTACRATTG TGTGCGGTGT GGGTGGCTC CAGGCCACG CTCCCTCTG 3480
 ATTTGATTTT TTGAGTTTTC GNTTGAAGA TCTACACAGN CAGTTACAC TGATCTTAA 3540
 GGACATATNT TATAACCCCTT TAAACAAAATCCCTC TCAATTCTAT TTGAGATGA 3600
 ATTCGATAC AGACTAGATG TCTTCTGAA GATCAATTAG ACATTTGAA ATGATTTA 3660
 AGTGTTTTCTT TTAATGTTCT CTGAANAAA STTTCCTTTG TATTTTABC CAAAAGTG 3720
 CCCTTTTGT CACIGGTTC TCTAGCATT CATGATTTT TTTCTCCTCA ATGAAATAAA 3780
 ATTCGATAAA TCACTGGACTG GCTTTCTGCT TGGATTTGAG GAAAGATGTG TTTAGGCCA 3840
 GAGCTTTCTT CAGTATTTGAA TTGTTTCCC CAGTATTTGA TTGTTTAAAT ATATACACAT 3900
 AGGAGCTGCA TTAAACACCT GCTGGTTAA ATTCGTCAT ATTCACCTC TACGTTTTA 3960
 GTATGCCRAA TCAATTTA CTTTACTTA AGCATTTGA ATTTGGAGTA TCTGGTACTA 4020
 GCTAAGAAAT AATTCATTA TTGAGTTTG TACTCNCRAA ATATGGGTCA TTCTCATGM 4080
 ATATGTMCC CCCAATGCG CTTCATTTTC CAGANACCTT GAGCCAGGAT ATTTTTTC 4140
 ATCAATTAGG TCCCCAAA 4159

<210> 5
 <211> 1708
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 5

AGGGACGCTG CGCGCACCGCC CCAAGTTAAC CCGGGGAGCC ATCATGAAAGC TGAATGCCA 60
 CCAGATGCAAG AACCAATGCCA TGAAGGTCTC CTACATCCCC GATGAGCAGA TAGCACAGG 120
 ACCTGAAAT CGGCGGCCAG GGGGCTTCTGG CTCTCGGGCT CAGCCUCCGCC ACGGCTCAGC 180
 TGTGGCAGCC GGGGCCCGAG CCAAGCAGCA GAAAGTGGAC ATCCUCCTTC CGCTCTGGT 240
 GCCCCACCCAG TATGTGGGIG CCATTATTC CAAAGGAGGC CGAACCATCC GCAACATCAC 300
 AAAACGACCC CAGTCACAGA TAGACCTGCA TAGGAGGGAC AACGAGGAG CAGCTGAAA 360
 ACCCATCAGT GTGCACTCC CCGCTGAGGG CTGCTCCTCC GCTTGTAAAGA TCAATTGGA 420
 GATTATGCAT AAAAAGGUTA AGGACACCA AACGGGCTGAC GAGGTTCCCCC TCAAGATCCT 480
 GGGCCATATAA TACTTGTAG GGGCTCTCAT TGGCRAAGGA GAAAGGAAAC TGAAAGAGGT 540
 AGACCAAGAT ATCCGRGACAA AAATCACCCT CTCTCGTTT CAGAACCTTA CCCTTACAA 600
 CCCTGAGAGG ACCNTACTG TGAAGGGCC CATCGAGAAT TGTGCAAGG CCGACCAAGGA 660
 AATAATGAAAG AAAGTTGGG AGGCTATGA GATGATGTC GCTGCCATGA GCTCTCACCT 720
 GATCCCTGGC CTGAACCTGG CTGCTGTAGG TCTTTTCCCA CTTCATCCA CGCCAGTCCC 780
 GCGCGCTCCC AGCAGCGTTA CTGGGGCTGC TCCCTATAGO TCCCTTATTC AGGCTCCCGA 840
 GCAAGGAGATG CTCCACCTGT TTATCCCCGC CCAGGCAGTG GGCCCCATCA TGGGCAAGAA 900
 GGGGCGACAC ATCCACACAGC TCTCCCGTT TGCCAGCGCC TCCATCAAGA TTGCAACACC 960
 CGAAACACCT GACTCCAAAG TTGCTATGGT TATCATCACT GGAACGGCCAG AGGCCCCATT 1020
 CAGCTCTAG GGAAGAAATCT ATGGCCTACCT CAAAGGAGGAG AACTTCTTTC GTCGCCAGGA 1080
 GGAAGTGAAG CTGGGAGACCC ACATACGTT GCAACCATCA GCAAGCTGGC GGGTCAATTGG 1140
 CAAGAGTCAA AAAACGGTCA ACCGACTTCA CAATTGAGG GCAAGCTAGG TGTAGTACCC 1200
 AAGAGACCGAG ACCCTCTGAG AGAACGACCA GGTCATCGT AAATCATCTG GACATTCTA 1260
 TCCCACTGAG ATGGCTCTAA CGAAGATCCG AGACATCTG CGCCAGGTTA AGCAAGCAGCA 1320
 TCAGAGGGCA CAGAGTACCC AGGCCCCAGGC AGGGAGGAAAG TGACCGAGGCC CTCCCTCTCC 1380
 CTTNGACTCC AGGACACCAA CGGGCAGAA TCGAGGAGTGT GCTCTCCCCG GCAGGCGTGA 1440
 GAAATGAGTCC GAAATCCGGGA CAATCGGGCC GGGGCTGTAGA TCAAGTTGTC CCAGCTTGAATT 1500
 GAGAAGAGATG TTCCAGTGAG GAAACCTGTAT CTNTLAGGCCA CAAACACCCA CCCAATGGC 1560
 CCAACACTGT NTGCCCCCTCG GGGTGTGAGA AATTTAGGG CAGGGCACTT TTAAGCTGG 1620
 ATTGTAAAT GAAAGCTCTCC AGGCCACACC AAGAGGGTGG ATCAACACCTC AGTGGGAGA 1680
 AAAATAAAAT TTCCCTCAGG TTTTAAA 1708

<210> 6
 <211> 3412
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 6

GGGCAGCGAGG GAGGCCGAGGA GGGCCGGGTÀ CCGGGCCCCGGG GGGAGCGCGGG GGTCTCGGGG 60
 AGAGAGACCGA TGATGAAACAA GCTTACATC GGGAAACCTGA GCCCCCGCGT GACCGCCGAC 120
 GACCTCGGGG AGCTCTTGGG GAGACAGGAAG CTGCCCCCTGG CAGGACAGAATG CCTGCTGAG 180
 TCCGGCTAAG CCTTCCTGGA CTAACCCGAC CAGAACTCGG CCATCCGGG CAGCAGAGCC 240
 CTCTCGGGTA AAGTGGAAATT GCATGGGAAA ATCAGTGGAAAG TTGATTTACTC AGTCCTAA 300
 AAGCTAAGGA GAGGGAAATAT TCGATTTGA AAGCTCCCTC CTCACCTGCA GTGGGAGGGT 360
 TTGGATGGAC TTTGGCTCA ATATGGGACA GTGGAGAAATC TGGAAAGT CAGACAGAC 420
 ACAGAAACCG CGTGTGCTAA CGTCACATAT GCAACAAAGAG AAGAAAGCAA ATAGAGCTG 480
 GAGAGACTAA CGGGGCACTA GTTGGAGAAC TACCTCTCA AGATTTCTCA CATCCCGGAT 540
 GAGGAGCTGA GTCOCCCTTC AGGCGCTCG AGAGCCCACTG GTGGGGAGCA CTCTTCCGG 600
 GAGCAGGGCC AGGCGCTGGG AGGCACTCT CAGGCCAGAC AGATTTGATTT CGCGCTGOGG 660
 ATCCCTGGTCC CGACCCACTT TGTGGTGGCC ATCATCGGAA AGGAGGGCTT GACCATAGG 720
 AACATCCTAA AGCAGACCCA GTCCCCGTTA GATACCTCTA GAAAAGAGGA CTCTGGAGCT 780
 CGAGGAGAGC CTGTCACCAT CCTATGCCACC CGAGAGGGGA CTTCTGAGAC ATGCGCGCATG 840
 ATTCTTGAAA TCTTGCAGAA AGAGGCAATG GAGACCAAAAC TAGGCCAGAGA GATTCCTCTG 900
 AAGATCTTGG GACACAATGG CTTGGTTGGA AAGTGTATTG GAAAGAACCG CAGAAATTG 960
 AAGGAAATTG AAGATGAAAC AGGGACCCAG ATAACAAATCT CATCTTGCAGA CGATTTGAGC 1020
 ATATACGAGC CGGAAGAGC CTCACATGTG AAGGGCGAGAG TTGGGGCTG TGCCACTCCT 1080
 GAGATAGAGA TTATGAAAGA GGTGCCTGAG GCTTITGAAA ATGATATGUT GGCTGTTAAC 1140
 CGACAGGCCA ATCTGATCCC AGGGTGTGAA CTCAGCGCAC TTGGCATCTT TTCAACACGA 1200
 CTGTGGTGC TATCTCCACC AGCAGGGCCC CGCGGAGGCTC CCCCCGGTGC CCCCCTACAC 1260
 CCUTCTACTA CCCACTCCGG ATACTCTCC AGCCTGTGACC CCTCATCCTA GTTTGGCCCG 1320
 TTCCCGCATC ATCACTCTTA TCCAGAGCAG CAGATGTGAA ATCTCTTCTAT CCCAACCCAC 1380
 GCTGTGGGCG CCATCATCGG GAGAAGGGG CCAACATCA AACAGCTGGC GAGATCTCCC 1440
 GGAGCCTCTA TCAAGATTCG CGTGGGGAA GGGCCAGAGC TCAACUAAAG GATGCTCATC 1500
 ATCAGUGGGC CACCCGAAAGC CGAGTCAAG GCCCAGGGAC GGATCTTGG GAAACTCAA 1560
 EGGGAAACT TCTTTAACCC CAAAGAAGAA CTGAAGCTGG AAGCGUATAT CAGACTGCC 1620
 TCTTCCACAG CTGGCCGGGT GATTGGCATA CCTGGCAGA CGTGTAGAGA ACTCCAGAAC 1680
 TTAACAGTG CGAAGGTCAAT CGTGGCTCTT GACCAACAGC CAGATGAAAA TGAGGAAGTG 1740
 ATGTCAGAA TTATCGGGCA CTTCTTGCCT AGCCAGACTG CACAGGCAA GATCAGGGAA 1800
 ATTGTACAAAC AGGTCAAGCA GCAAGGAGCAG AAATACCCCTC AGGEGATGCG CTCACAGC 1860
 AGCAGTGAAG GCTCCCACAG GAAAGAGCAA AACAAAGGAAAT GAAATGAGCC CTTCCAAACAC 1920
 CTGACAGGAA GAGACCAAAC GCAAGCGAGCC AGATCGGGAG CAAACCAAAG ACCATCTGAG 1980
 GATGAGAGG TCTGGGGAGG CGGAGGAGGA CTCTCCCGAG GCCCCGAGAA CCCCAGGGCC 2040
 CGAGGAGGGG CGGGGAAGGT CAGCGAGGTT TCCCGAAACCC ACCGGRGCCCC GCCTCCCGCC 2100
 CCCCCAGGGCT TCTGCGAGGCT TCAGCCATCC ACTTCACCAT CCTACTGGAT CTCTCTGAA 2160
 CTCCGACGAA GCTATCCCTT TTGATGAAAC TAACTATAGGT AAAGCTGTTT AAAGCCAAAGC 2220
 AATATTCAGCA CCTTTTTCTT GTGCGAAATC CTCTCTCTAC ATGTGTUTAC ATATTAAGAAA 2280
 GGGAGATGT TAAAGTATGT GGCCTGGG TTACACAGGG TGOCTGAGC GGTAAATATAT 2340
 TTTAGAAGATA ATATATCAAAT TAACTCAACT AAGTCCAACT TTTAAATCAAT TATTAATTTT 2400
 TTTTTCTTTT TAAAGAGAAA CGAGGCTTT CTAGACTTTA MAGAATAAG TCTTTCGGAG 2460
 CTCTCACGGT GTAGAGAGGA GCTTGGGGC CACCCGACAA AAATTCAACU AGAGGCAAAT 2520
 CTCCCTGGAA GGACACTCAC GGCAGTTCTG GATCACCCCT GTATGTCAAC AGPAGGGATA 2580
 CCTCTCTCTT GAGGAGGAA CTCTGCACT CCTCATGCGCT CTCTAGCTCA TACACCCATT 2640
 TCTCTTGTCT TCACAGGTT TAAACTGGTT TTTGCATAC TCCATATATA TTCTCTGTCT 2700
 CTCTCTGTCTT ATCTCTCCCCC TCCCTCCCCC CCTCTCTTC TCCATCTCA TTCTTITGAA 2760
 TTTCCTCATC CCTCCATCTC AATCCCGTAT CTAGGACCCC CCCCCCCCCC AGGCAAAAGCA 2820
 GTGCTCTGAG TATCACATCA CACAAAGGA ACAGAAGCGA AACACACAAA CGAGGCTCAA 2880
 CTTACACTTG GTTACTCAAA AGAACAGAGC TCAATGGTAC TTGCTCTAGC GTTTGGAG 2940
 AGGAAACAG GAAACCCACCA AAACGACCAA TCAACCCAAAC AAAGAAGAAA TTCCACAAATG 3000
 AAAGAATGTA TTTCGTCTT TTGATTTTCGCTATGAGC CATCAATATT CGGCAAAATG 3060
 ATTCTTTCTT TTAAAGAAA AATGTGGAG CAAAGTNGAA ATTTACCAAG GTTGGTGGCC 3120
 CAGGGCGTTA AATTCACAGA TTTTTTAAC CAGAAAGACA CACAGAGAA GCTACCTCAG 3180
 GTGTTTTTAC CTCAGCCTAC TGCTCTTGT TTCCCTTAG AGATTTCTA AGCTGATAG 3240
 TTGGAGCATT TTTTTATTTT TTAAATAAA ATGAGTNGAA AAAAAAATAA GATATCAAC 3300
 GCCAGCCTGG AGAAGGTGAC AGTCCAGTG TGCACAGCT GTCCTGAATT GTCCTCGGCT 3360
 AGCCAGAGAAC CNATATGGCC TTCTTTGGA CAAACCTTGA AAATCTTAT TT 3412

<210> 7
 <211> 1946
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 7

GCTGTAGCGG AGGGGGCTGGG GGGCTGCTCT GTCCCGCTTC TTGGCCGCTTA CGGGCTCAGC 60
 CCACCCAGAG GCCTGGGTGG GAGGCGAGT GCTCAGCTTC CGGGCTTAAA AGGGGGGAAA 120
 TTCAAAATCCG AATATATCCCA CCCAGCTCC GATGGGAAGT ACTGGACAGC CTGGCTGGCTC 180
 AGTATGGTAC AGTAGTGAGAC TGTGAGCAAG TGAACACCGA GAGTCAGACG GCACTGGTGA 240
 ATGTCACCTA TTCCACCGG GAGCAGACCA GGCAGAGCCAT CATGAGGCTG AATGGCCACC 300
 AGTTGGAGAA CCTGTGGCTTG AAGGTCTCT ACATCCCGA TGGAGCAGATA GCACAGGAC 360
 CTGACAATGG GTGGCCGAGGG GGCTTTGGCT CTGGGGCTCA GCGGGGGAGG GGETCAGCTG 420
 TGGCAGCCGG GGCAGGAGCC AAGCAGCAGC AAGTGGACAT CGGGCTTCGG CTGGCTGGTGC 480
 CCACCCAGTA TGTGGGTGCC ATTATTCGCA AGGGAGGGGG CACCATCCCC AACATCAGAA 540
 AACAGACCCA GTCCAGATA GAGGTGCTA GGAAGGAGAA CGCAGGTGCA GCTGAAAAG 600
 CCATCAGTGT GCACTCCACG CCTGAGGGCT GCTCCTCCGC TTGTAAGATG ATCTTGGAGA 660
 TTATGCATAA AGAGGCTAAG GTCACCAAAA CGGCTGACGA GGTTCGGCTG AAGATCCTGC 720
 CCCATAATAA CTTTGTAGGG UGTCTCATTC GCAAGGAAGG ACGGGACCTG AAGGAUGTAG 780
 AGCAAGATAA CGAGACAAAATCAGACAT CTCCTGTTGCA AGACGTTAACG CTTTACACCC 840
 UTGAGAGGAC CATCACTGTG AAGGGGGCCA TCGAGAAATTG TTGCAGGGGC GACCCAGAA 900
 TAATGAAGAA AGTTGGGGAGG GCGCTATGAGA ATGATCTGTC TGCATGAGC TCTCACCTGA 960
 TCCCTGGUCT GAACTGGTGT GCTGAGGTC TTTCTCCAGC TTCACTCCAGG CGAGTCCCCG 1020
 CGCCCTCCAG CAGCGTTACT GGGGCTGCTC CCTATAGCTC CTTTATGAG GCTCCCGAGC 1080
 AGGAGATGGT GCAAGGTGTTT ATGCCCGGCC AGGCACTGGG CGGCACTCATG GGCAGAAGG 1140
 GGCAGCACAT CAAACAGCTC TOOCGGTTTG CGAGGCCCTC CATCAAGATG GCACCAACCG 1200
 AAACACCTGA CTCCAAAGTT CGTATGGTTA TCATCACTGG ACGGCCAGAG GCGGCAATTCA 1260
 AGGCCTCAGGG AAGAAATCTAT GGCACACTCA AGGAGGAGAA CTTCTTGGT CGCAAGGAGG 1320
 AAGTGAAAGCT CGAGACCCAC ATACGTGTGC CAGCACACAGC AGCTGGGUGG GTCATGGCA 1380
 AAGGTGGAAA AACCGGTGAAC GAGTGGCAGA ATTGACGGC AGCTGAGGTG CTAGTACCAA 1440
 GAGACCAAGAC CGCTGATGAG AACGAOCAGG TCATOGTCAA AATCACTGGG CATTCTATG 1500
 CCPAGTCAGAT GGCTCAGCG AAGATOCGAG ACATCCCTGGC CGAGCTTAAG CAGCAGCTC 1560
 AGAAGGGACA GAGTGTACCG GCCCAGGCAC GGAGGAGTGC ACCAGGCGCT CCCTGCCCCP 1620
 TNGAGTCCAG GACACACAGC GGCAGAAATC GAGAGTGTGC TTGCTCCGGC AGGGCGCAGA 1680
 ATGAGTGGGA ATCCGGGACA CNTGGGGCGG GCGTGTAGATC AGGTGGTGC ACTTGATTGA 1740
 GAAAGGGT CCACTGAGGA AGCCCTGATCT NTGAGCCCCA AACACCCACG CAATTGGCCC 1800
 AACACTGTNT GCGCCCTCGG GTGTCAGAAAT TTNTAGGGCA AGGCACTTTT AAACCTGGAT 1860
 TGYTTAAGA AGCTCTCCAG GCGGCAACCA GAGGGTGGAT CACACCTGAG TGGGAAGAAA 1920
 AATAAAATTT CCTTCAGGTT TAAAAA 1946

<210> 8
 <211> 3293
 <212> DNA
 <213> Homo sapiens
 <220>
 <400> 8

GGCAGCCGAG GAGGGCGAGGA GCGCCGGGTA CGGGGGCGGGG GGAGGCGGGG CCTCTCCGGG 60
 AAGAGACGGA TGATGAAACAA CCTTTACATC CGGGACCTGA CGCCCTGGCGT CACCCCGCAGC 120
 GACCTCCGGC AGCTCTTG GGCAGGGAG CTGGCCCGTG CGGGAACAGGT CCTGCTGAAC 180
 TCGGGCTACG CCTTCGTGCA CTACCCCGAC CAGAACTGGG CCATCCGGCGC CATCGAGACC 240
 CTCTGGGTAA AGTGGAAATT CGATGGGAAATC ATCATGGAAAG TTGATTAATC AGCTCTCTAA 300
 AACCTTGGGA CGAGGAAATT TCAGATTGCA AACATOCCTC CTACACTGCA GTGGGAGGTG 360
 TTGGATGGAC TTGGGCTCA ATATGGGACA GTGGAGAATG TGGAAACAGT CAACACAGAC 420
 ACAGAAACCG CGGTTGTCAA CGTCACATAT GCAACAGAG AGAAGCAGA ATAGCCATG 480
 GAGAAGCTAA CGGGGCATCA GTTTCGAGAC TACTCTTCA AGATTTCTA CATCGCGAT 540
 GAGAGGGTGA GCTCCCTTC GCGGCGCTCAG CGAGGCCAGC GTGGGGACCA CTCTTCCGGG 600
 GAGGAGGGCC AGCGCCCTGG GGGCACTTCT CAGGCCAGAC AGATGGATT CGCGCTGGGG 660
 ATCCTGGTCC CCACTTCAAGT TTGGGTTGCA ATCATGGAA AGGAGGGCTT GACCATTAAG 720
 AACATCACTA AGCAACACCA GTCACGGGTA GATATCCATA GAAAAGACAA CTCTGGAGCT 780
 CGAGACAGC CTGTCACCAT CCATGCCACC CGAGGGGGG CTTCTGAAAC ATGCCCGATG 840
 ATTCTTGAAA TCATGCAAGA AGAGGCAAGAT GAGACCAACG TAGCCGAAGA GATTCCTCTG 900
 AAAATCTTGG CACACAAATGG CTGGGTGGA AGACTGATTC GAAAAGAAGG CAGAAATTG 960

AAGAAATTG AACATGAAAC AGGGACCAAG ATAACAATCT CACCTTGCA GGATTTGAGC 1020
 ATATACRACC CGGAAAGAAC CATCACTGTG AAGGGCACAG TTGAGGCGTG TGCAGTGCT 1080
 GAGATAGAGA TTATGAGAA GCTGCGTGAG CCCTTTGAAAT ATGATATGCT GGCTGTTAAC 1140
 ACCCACTCCG GATACTTCU CAGCCTGTAC CCCCATCACC AGTTTGGCCC GTTCCCCCAT 1200
 CATCACTCTT ATTCAGAGCA CGAGATTGTG ATCTCTTCA TCCCACGCCA GGCTGTGGC 1260
 GCACATCATCG CGAGAAGGGG GGCACACATC AACACAGCTGG CGAGATTGCG CGAGAGCTCT 1320
 ATCAACATG CGCTGTGCGA AGGUCACAGAC GTCAGCGAAAGG GGTGTTGCTAT CATCACCGGG 1380
 CCACCCGAGG CGGAGTCAA GGGCAGGGG CGGATCTTGG GGAAGCTGAA AGAGGAAAC 1440
 TTCTTTAACG CTAAGAGAGA AGTGAAGCTG GAAGGGCATA TCAAGAGTCC CTCCTCCACA 1500
 GCTCCCCGGG TTTTGTGCGA AGGTCGGCAAG ACCCTGAACG AACCTGAGAA CTTTAACTACT 1560
 GCAGAAGTCA TCGTGCCTCG TGACCAAACG CGAGATGAAA ATGAGGAAGT GATGTCAGA 1620
 ATATCGGSC ATCTCTTTCG TACCGAGACT GCACAGCGCA AGATCAAGGA ATATGTACAA 1680
 CAGGTGPGAC AGCAGGAGCA GAATACCCCT CAGGGAGTCG CCTCACAGCG CAGCAAGTCA 1740
 GGCTCCCCACA GGCRCAGCA AAACAAACGGA TCAATGTAGC CCTTCAACA CCTGACAGAA 1800
 TGAGACCAAAGC CAGATCGGGG CCAACCCAAA GACCACATGCA GGAATGAGAA 1860
 GTCTCCGGAG GCGGGTGGGG ACTCTGCCG GGCCTTGAGA AGGCGAGGGG CGAGGGAGGG 1920
 GCGGGGAAGG TCAGCCGGT TTGCGAGAAC CACCGAGCCG CGCCTCCCCC CCCCCAGGGC 1980
 TTCTGCAGGC TTCAGCCATC CACTCAGCA TCCACTCGGA TCTCTGCTGA ACTCCCCACGA 2040
 CGCTATCCCT TTTAGTTGAA CTAACTTGG TCAACCTGCT TCAAGCCAGG CAAATGCAAC 2100
 ACGCTTTTTC TCTGGCAAT CGTCTCTGTA CATGTGTGTA CATATTAAGA AGGGAAAGCTG 2160
 TTATGATATG TGGCCTGTGTT GTTACACAGG GTGCTGTGAG CGGTAAATATA TTATGAAAT 2220
 ATATATATCAA ATAATCTAAC TRACTCCAAT TTATTAATCAA TTATTAATTT TTCTTCTTT 2280
 TTAAAGAGAA AGCAGGGCCTT TCTAGACTT AAAGATTAAG GTCTTTGGGA GGTCTGAEGG 2340
 TTAGAGAGG AGCTTTGAGG CCACCCGAC AACATTACCC CAGGGAAA TCTCGTCGGG 2400
 AGTCACACTC CGGGAGTCTT GGATCACCTG TGATGTGCA CAGAAGGGAT AGCGTCTCT 2460
 TGAGGAGGA ACTCTCTCAC TCTCATGCC TGCTCTAGCTC AACACCCAT TTCTCTTTGC 2520
 TTCAACGGTT TTPACCTGGT TTCTGCATA CTGCTATATA ATTCTCTGTC TTCTCTGTT 2580
 TATCTCTCCC CTCCCTCCCC TCCCTCTCTT CTCATCTCC ATTCTTTGAA ATTCTCTCAT 2640
 CCCTCCATCT CAATCCCGTA TCTACGCCG CCCCCCCCCCCC CAGGCAAAGC AETGCTCTGA 2700
 GTATCACATC ACACAAAAGG AACAAAGG AAACACACAA ACCAGCTCTCA ACTTAUACIT 2760
 GGTACTCAA AACAAACAAGA GTCACTGGTA CTGTCCTAG CGTTTTGGAA GGGAAAACA 2820
 GGAAACCCACC AACCCACCA ATCACCAA CAAAGAAAAA ATTCCACRAT GAAAGAATGT 2880
 ATTTTGTCTT TTGCAATTG GGTGTATAAG CTCATCTATAT TCAGCRAAAT GATTCCCTTC 2940
 TTATTAAGAAA AACATCTGGA CGAAGATGAA ATTATACCAA GGTGTTGTCG CCAAGGGCGTT 3000
 AAATTCACAG ATTTTTTAA CGAGAAAC ACACAGAGA AGCTACCTCA GGTGTTTTTA 3060
 CCTCAGGCC TTGCTCTGT GTTTCCTTA GAGATTTCTC AACAGCTGATA GTTGGAGCAT 3120
 TTCTTATTT TTATTAATAA ATGAGTTGG AAAAAGAAATA AGATATCAG TCCAGGCTG 3180
 GAGAAGGTGA CAGTCCAAGT GTGCAACAGC TGTCTGAAT TGTTTCCCG TAGCCAAGAA 3240
 CCNATATGGC CTTCTTTGG AACAAACCTG AAAATGTTA TTT 3283

